

Interoperability of federated, distributed information architectures

(Mostly) conceptual, (some) technical and (a few) political aspects of a core notion on the verge of degenerating into a buzzword

Prof. Dr. Stefan Gradmann
Humboldt-Universität zu Berlin / School of Library and Information Science
stefan.gradmann@ibi.hu-berlin.de

The Funny Talk You Won't Get ...



Overview



■ Intro

- Some politics (EC FP7, Europeana)
- Intended Audience, Definitions, Motivations

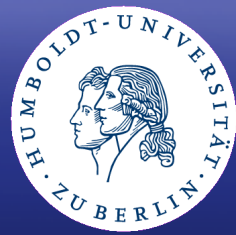
■ Technical

- Interoperability aspects of **Selected Frameworks** for DL modeling
 - DELOS, 5S, DRIVER, OAI-ORE, DCMI abstract, JISC Information Environment, JCR, iRODS
 - Thanks to Donatella Castelli, Ed Fox, Wolfram Horstmann, Andy Powell, Herbert van de Sompel, Pete Johnston and a lot more ...
 - Deliberately discarded: DAREnet, aDORe, CORDRA / IMS DRI (CP and ECL), e-Framework, O.K.I. Open Service Interface Definitions (OSIDs), and many more ...
- Six dimensions of the **interoperability matrix** abstracted from these frameworks plus some thoughts on **abstraction levels**
- Politics revisited: **Interoperability 2.0**

Background

- Google Books & related political trouble helped to trigger
- **EC i2010 (Lisbon) agenda** with Digital Libraries as one of 3 'flagship initiatives': the setting up of the European Digital Library as a common multilingual access point to Europe's distributed digital cultural heritage including all types of cultural heritage institutions
 - 2008: at least 2 million digital objects; multilingual; searchable and usable; work towards including archives.
 - 2010: at least 6 million digital objects; including also museums and private initiatives.
 - "I am not suggesting that the Commission creates a single library. I envisage a network of many digital libraries – in different institutions, across Europe." V. Reding (29 September 2005)
- => High level group, Expert group, **Interoperability group**
 - Contribute to the **short term DL agenda** => identify areas for short term action and recommend elements of an action plan (**list of prioritised feasible options**)
 - Contribute to the **long term DL agenda** => identify key elements for a **long term strategy**

Interoperability Group Composition

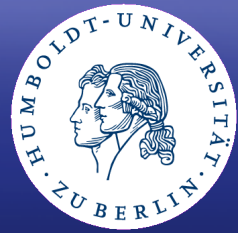


- Emmanuelle **Bermes** (Bibliothèque nationale de France / F)
- Mathieu **Le Brun** (Centre Virtuel de la Connaissance sur l'Europe / LU)
- Sally **Chambers** (The European Library Office / TEL),
- Robina **Clayphan** (The British Library / GB),
- Birte **Christensen-Dalsgaard** (State and University Library Aarhus / DK),
- David **Dawson** (The Museums, Libraries and Archives Council / GB),
- *Stefan **Gradmann** (Hamburg University Computing Center / D, Moderator),*
- Stefanos **Kollias** (Technical University of Athens / GR),
- Maria Luisa **Sanchez** (Ministerio de Cultura / ES),
- Guus **Schreiber** (Vrije Universiteit Amsterdam / NL),
- Olivier **de Solan** (Direction des Archives de France / F)
- Theo **van Veen** (Koninklijke Bibliotheek / NL)
- *EC: Pat **Manson** (Chair) and Marius **Snyders** (European Commission, DG INFSO, both Cultural Heritage and Technology Enhanced Learning) Federico **Milani** (European Commission, DG INFSO, eContentPlus)*

Definitions, definitions ...

- “Interoperability is the capability to **communicate, execute programs, or transfer data** among various **functional units** in a manner that requires the user to have **little or no knowledge** of the **unique characteristics** of those units.”
(ISO/IEC 2382 Information Technology Vocabulary)
- „the ability of two or more systems or components to **exchange information** and to **use** the information that has been exchanged.“
(IEEE)
- Interoperability is a property referring to the ability of diverse systems and organizations to work together (inter-operate). The term is often used in a technical systems engineering sense, **or alternatively in a broad sense, taking into account social, political, and organizational factors that impact system to system performance** (Wikipedia).
- => Plethora of definitions should make you suspicious!

Talk Motivation & Intended Audience



- “Open and interoperable are two words in the Information Technology world susceptible to **misunderstanding** at best, at worst to **self-serving abuse**. It is important to clarify their accepted meanings, because how they are understood in the market has direct practical consequences for consumers, vendors and regulatory authorities.” (European Committee for Interoperable Systems/ECIS, <http://www.interoperability.eu/>)
- Intended audience are institutions implied in “federated, distributed information architectures”:
 - Cultural heritage
 - Collective memory
 - Scientific and scholarly communities
- ... at senior technical or decision making level.

Interoperability Motivations: Europeana as Example



- Europeana will be federating objects from **distributed sources**
- Europeana will be federating objects from **heterogeneous sources** with **different community background** – e. g. libraries vs. museums vs. archives ... but also scholars vs. policy makers vs. meta users ...
- Europeana will be part of a bigger framework of **interacting global information networks** including e. g. 'Digital libraries', scientific repositories and commercial providers
- Europeana will have to be built with **minimal development efforts** and thus rely as much as possible on **web and internet standards** and **existing building blocks**
- And this is why interoperability figures so prominently place in the name of the “technical” WP of EDLnet: **Interoperability is the heart of the technical vision of Europeana!**

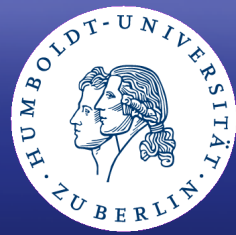
Inter-what?

Selected Frameworks
of Information Systems Architecture and Interoperation

DELOS

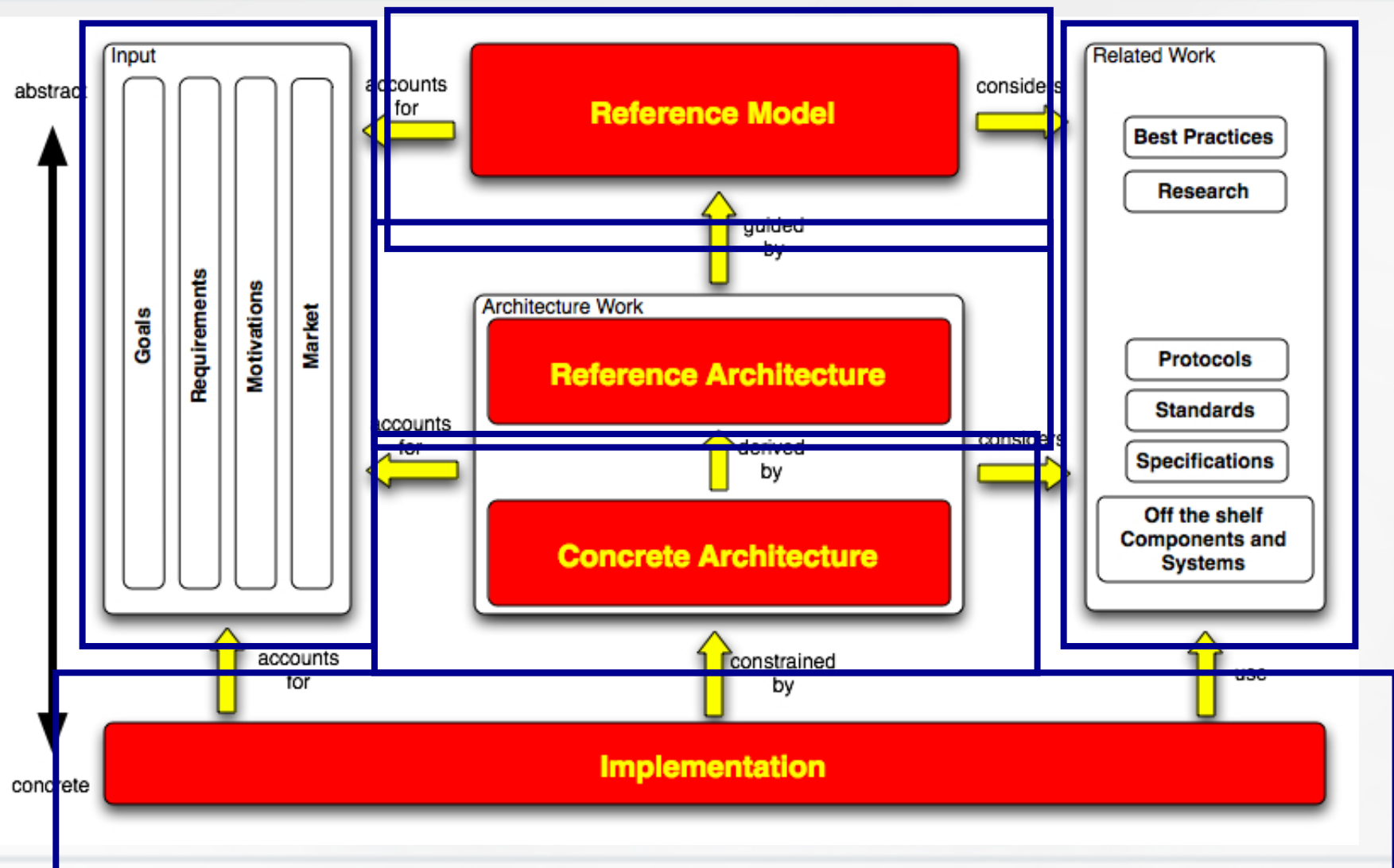
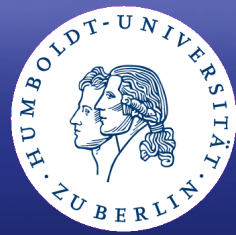
A Computer Science Based Framework

DELOS Reference model for DLMSs

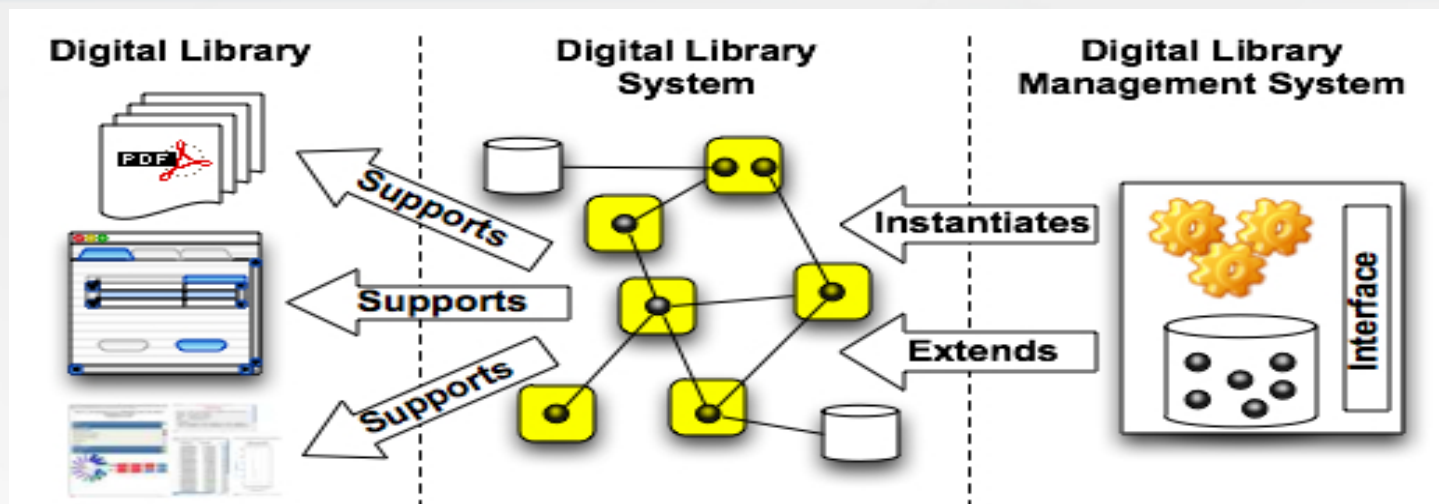


- Created by the **DELOS NoE**
- **The Digital Library Manifesto**
Objective: To set the foundations and identify the cornerstone concepts within the universe of Digital Libraries, facilitating the integration of research and proposing better ways of developing appropriate systems.
- **The Digital Library Reference Model**
Objective: To identify a number of concepts and relationships that represent the significant aspects of the different type of DL “systems”
- **A Digital Library Reference Architecture (unpublished draft!)**
Objective: To establish architectural guidelines for DL Systems supporting federated DLs

Reference Model for DLMs: Components and Relations

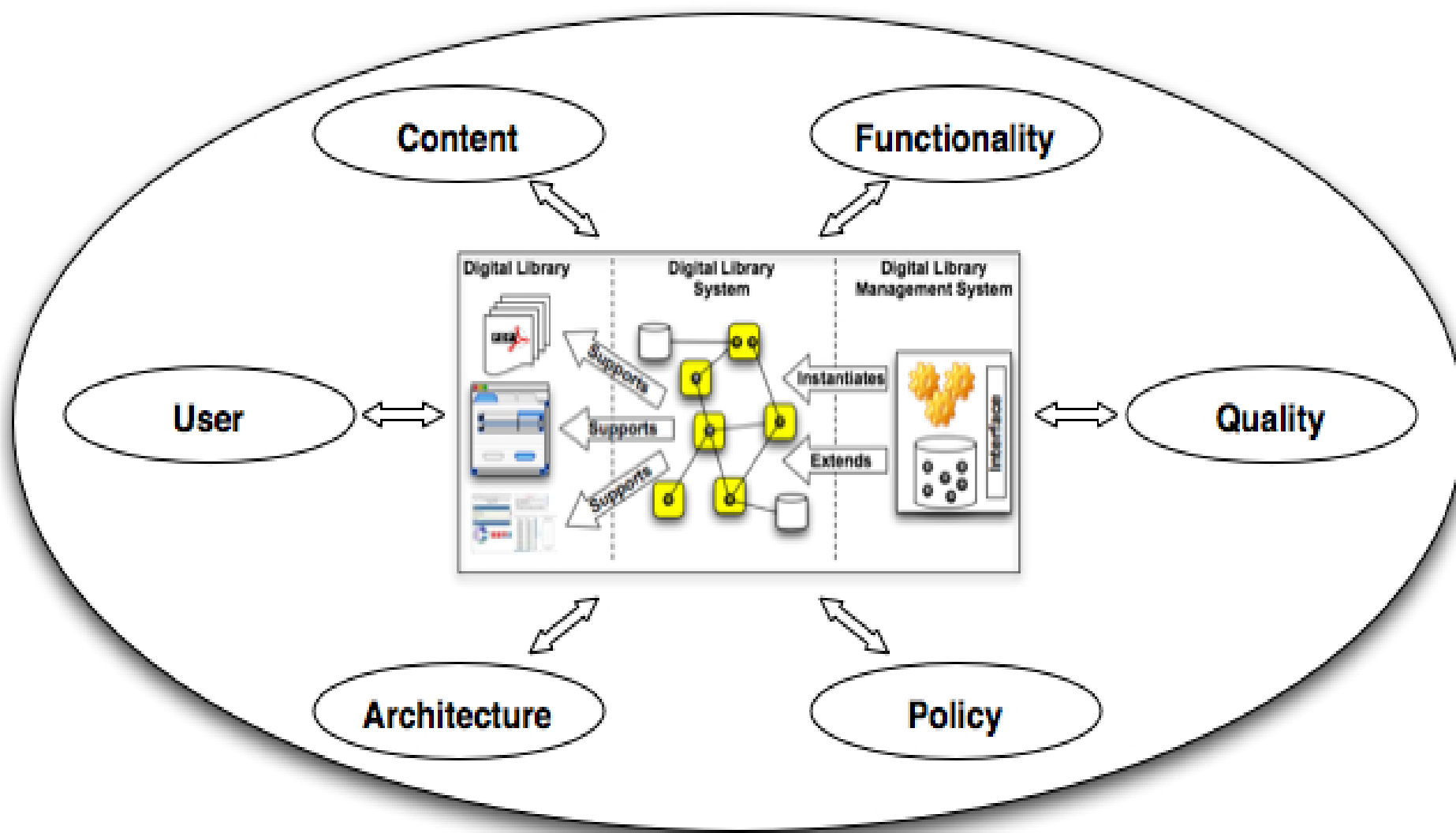


DL Manifesto: the “systems”

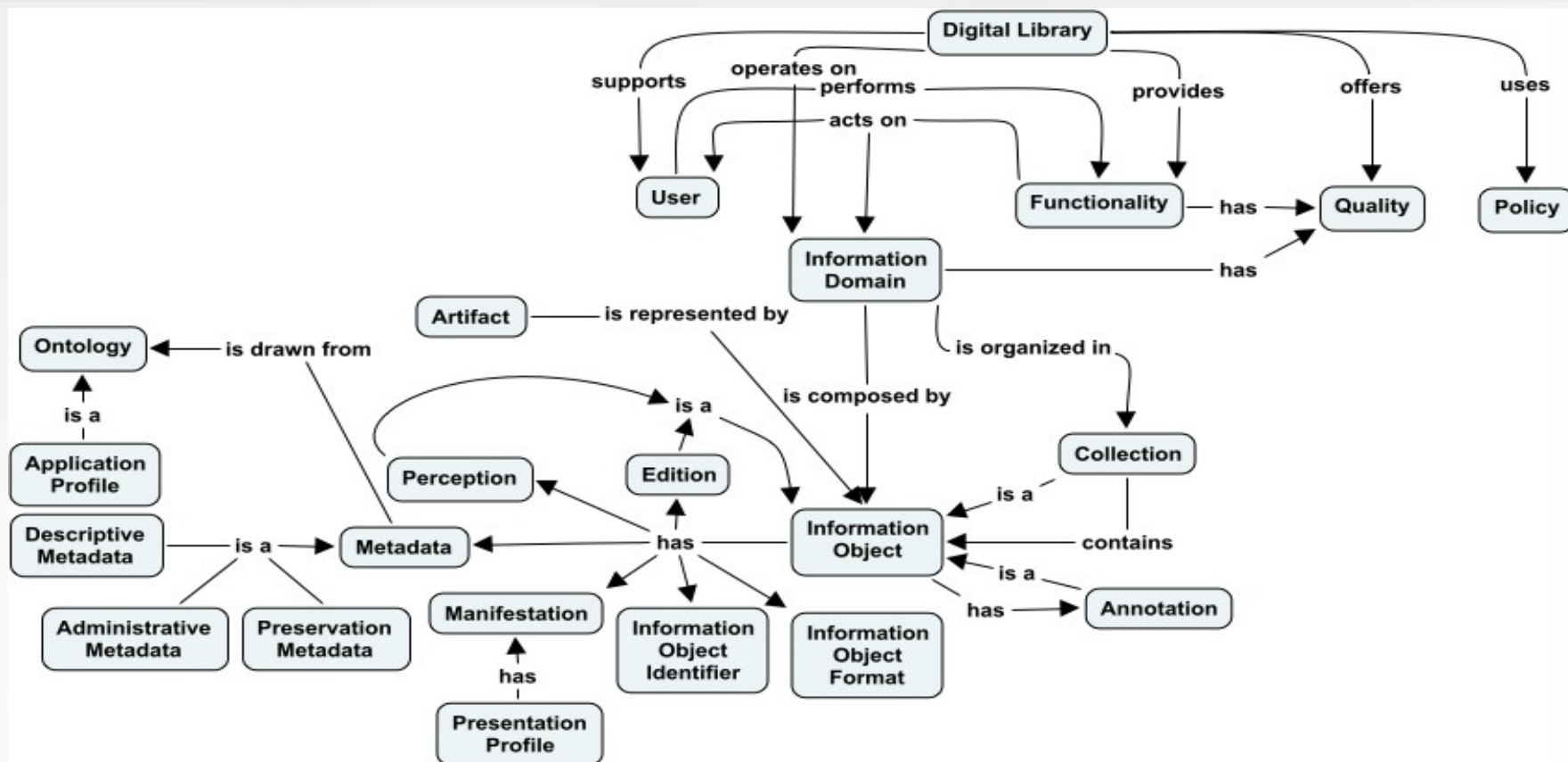


- **DL:** A (potentially virtual) **organization** that comprehensively collects, manages, and preserves for the long term rich **digital content** and offers to its user communities specialized **functionality** on that content, of measurable **quality**, and according to prescribed **policies**.
- **DLS:** A **software system** that is based on a (potentially distributed) architecture and **provides all functionality** that is required by a particular Digital Library. Users interact with a Digital Library through the corresponding Digital Library System.
- **DLMS:** A **generic software system** that provides the appropriate software infrastructure to both (i) produce and administer a Digital Library System that incorporates all functionality that is considered foundational for Digital Libraries and (ii) integrate additional software offering more refined, specialized, or advanced functionality (~ a **DLS factory**)

DL Reference Model



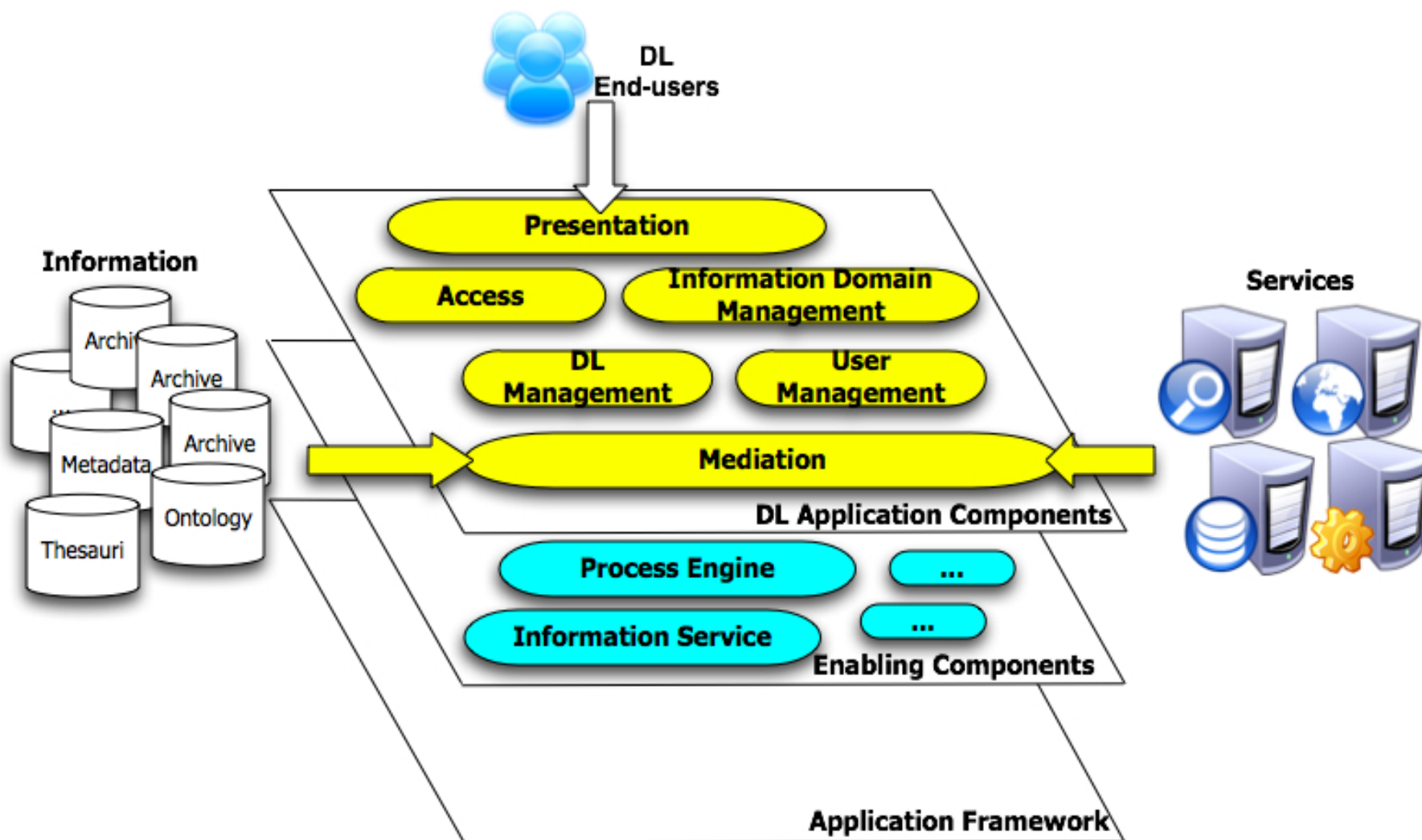
DL Reference Model: A High Level Concept Map



Definition 17. Information Object

*The main unit of information which is managed by the DL. An information object is a **DL Resource** identified by an **Information Object Identifier** and has associated **Metadata** for various management purposes. Moreover, an Information Object **MAY** have multiple **Editions**, **Perceptions** and **Manifestations**; and **MAY** be associated with **Annotations**.*

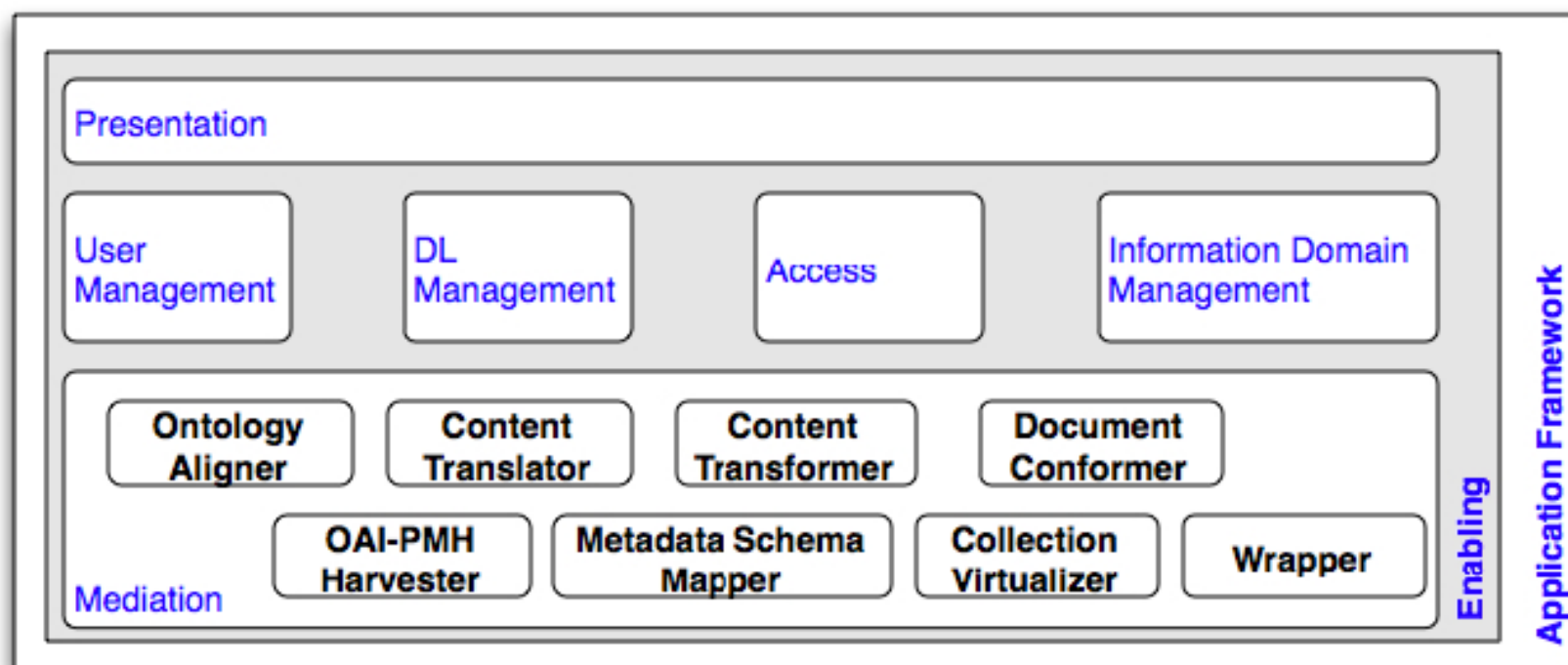
DL Reference Architecture: Functional Areas



DL Reference Architecture: Mediation Area



„Re-use, integration, interoperability are key requirements in the DL application area. In the current situation where no established rules nor principles exist for the development of digital library systems (DLS) the satisfaction of these requirements is difficult and has to be done on a case-by-case basis. In order to start overcoming this lack some of the DELOS partners active on the design of DL architectures decided to specify a *Reference Architecture* for component-based DLSs. These systems are particularly suitable to support one of the most important class of DLs: the *federated digital libraries*.“



DELOS Reference Framework: Characteristics



- Very abstract model rooted in Computer Science and only loosely related to cultural institutions' reality
- Although intended to create and enhance interoperability, the reference architecture still remains too abstract to really help
- It is unclear when work on the reference architecture will be taken up again and by whom
- The reference model is a very good starting point for conceptual work, even though it is not yet entirely mature and stable (and probably never will be, anyway!)

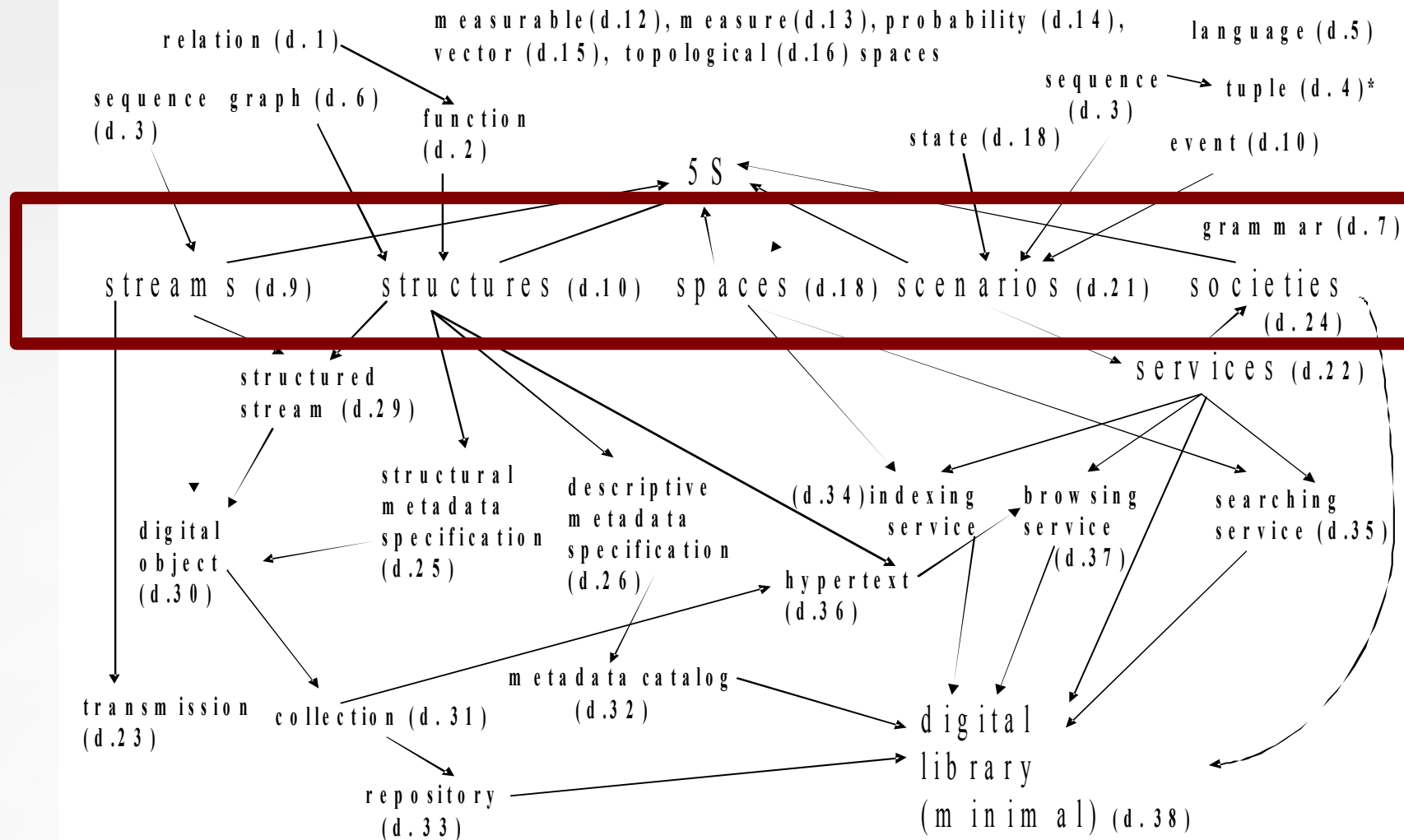
5S

Another Computer Science Based Framework

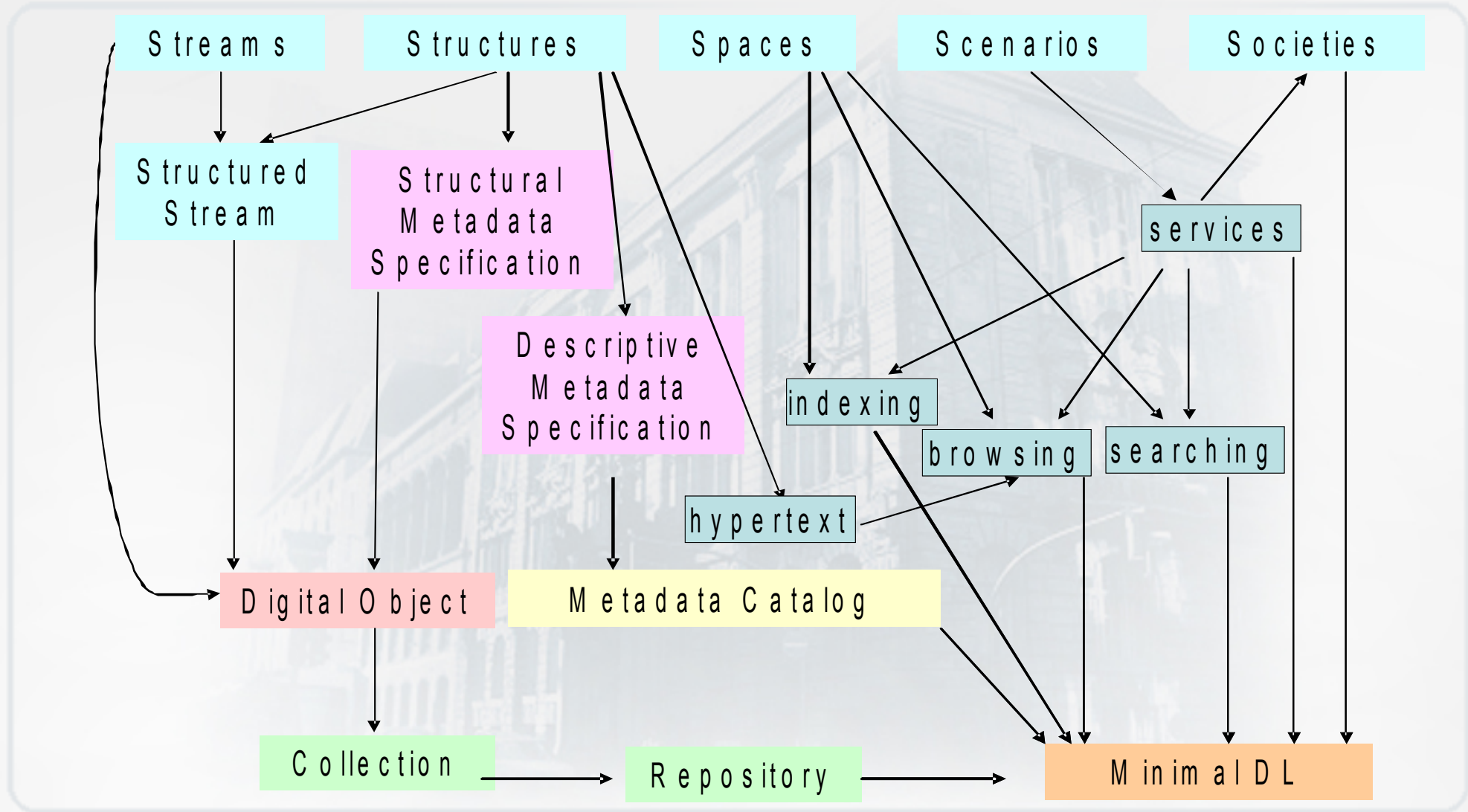
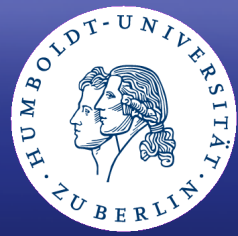
5S Model



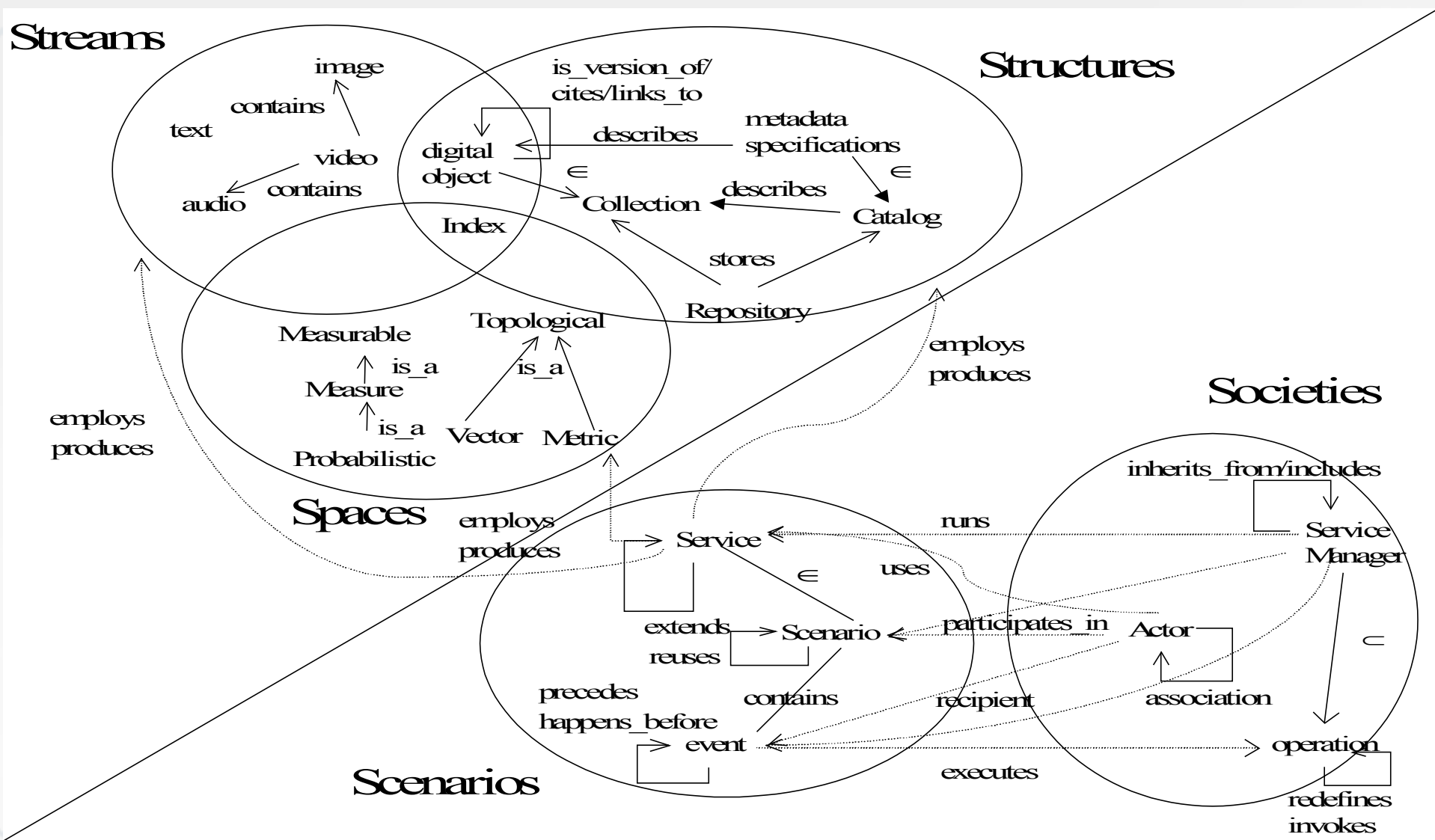
- Streams, Structures, Spaces, Scenarios and Societies (5S), is a unified formal theory for Digital Libraries (DLs) created within Ed Fox' group at Virginiatech.
- With 5S, digital library abstractions such as digital objects, metadata, collections, services, etc., can be rigorously and usefully described through compositions of basic and higher level mathematical objects.
- 5SL, an XML realization of the 5S model, is a domain-specific, declarative language for specifying and generating Digital Library applications.
- 5SL, 5S based ontologies and 5S metamodels together can be used to integrate individual DLs into a 'Union DL'



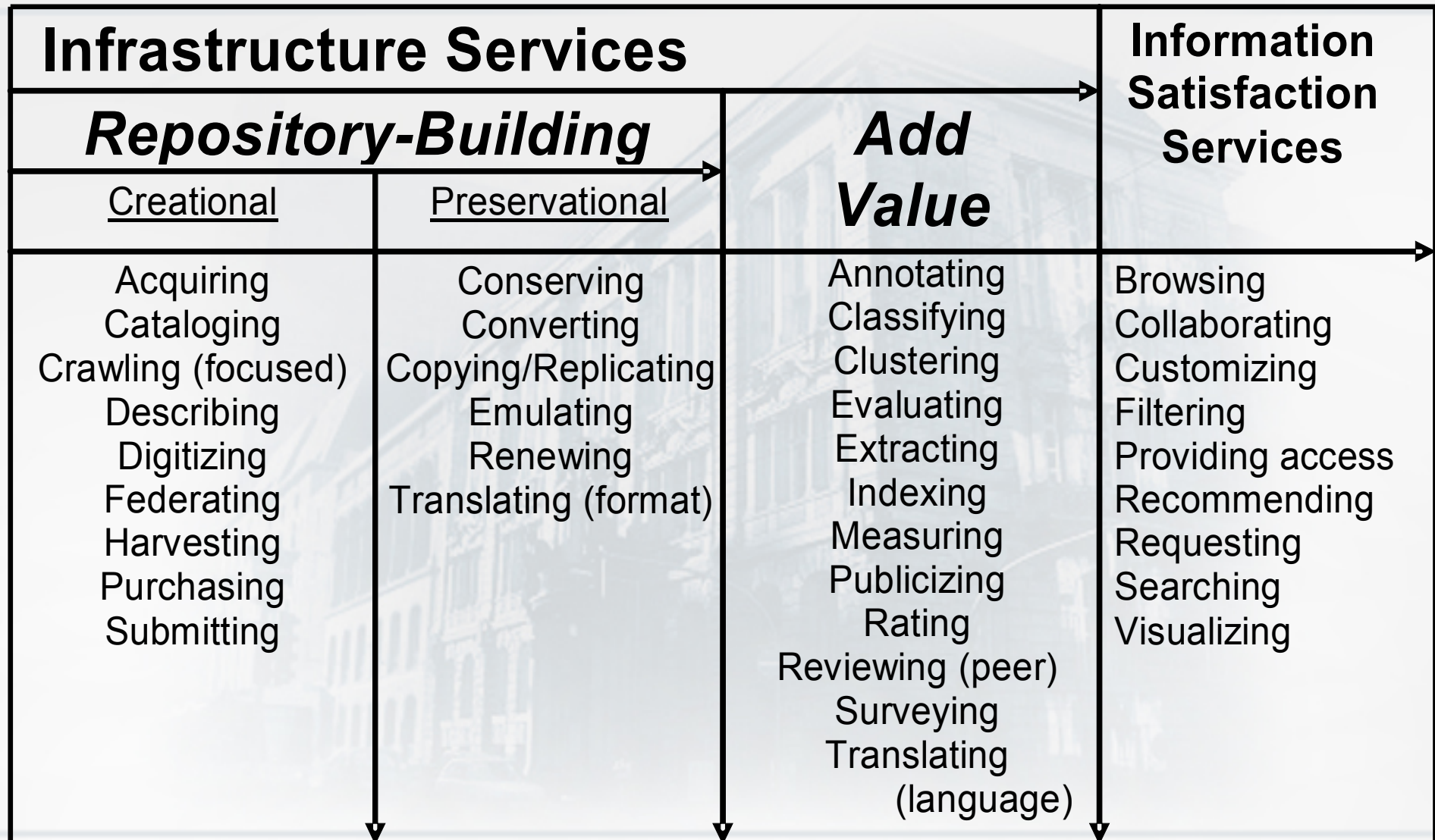
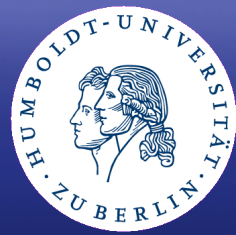
A minimal DL in the 5S Framework



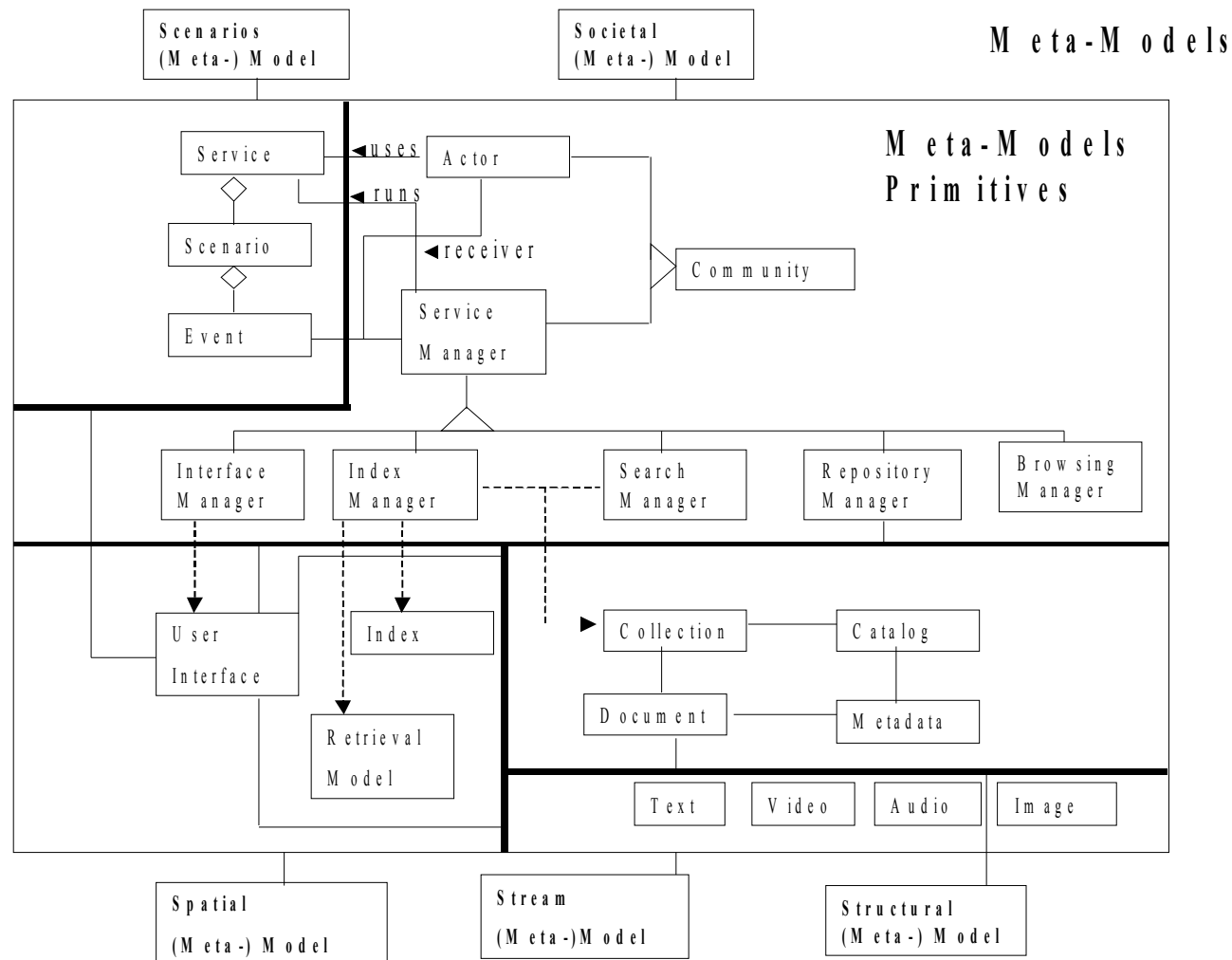
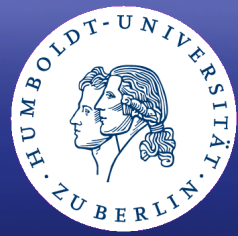
5S Ontology Domain: Relations View



5S Ontology Domain: Classified View



5S Metamodel Proposed for Conceptual Integration

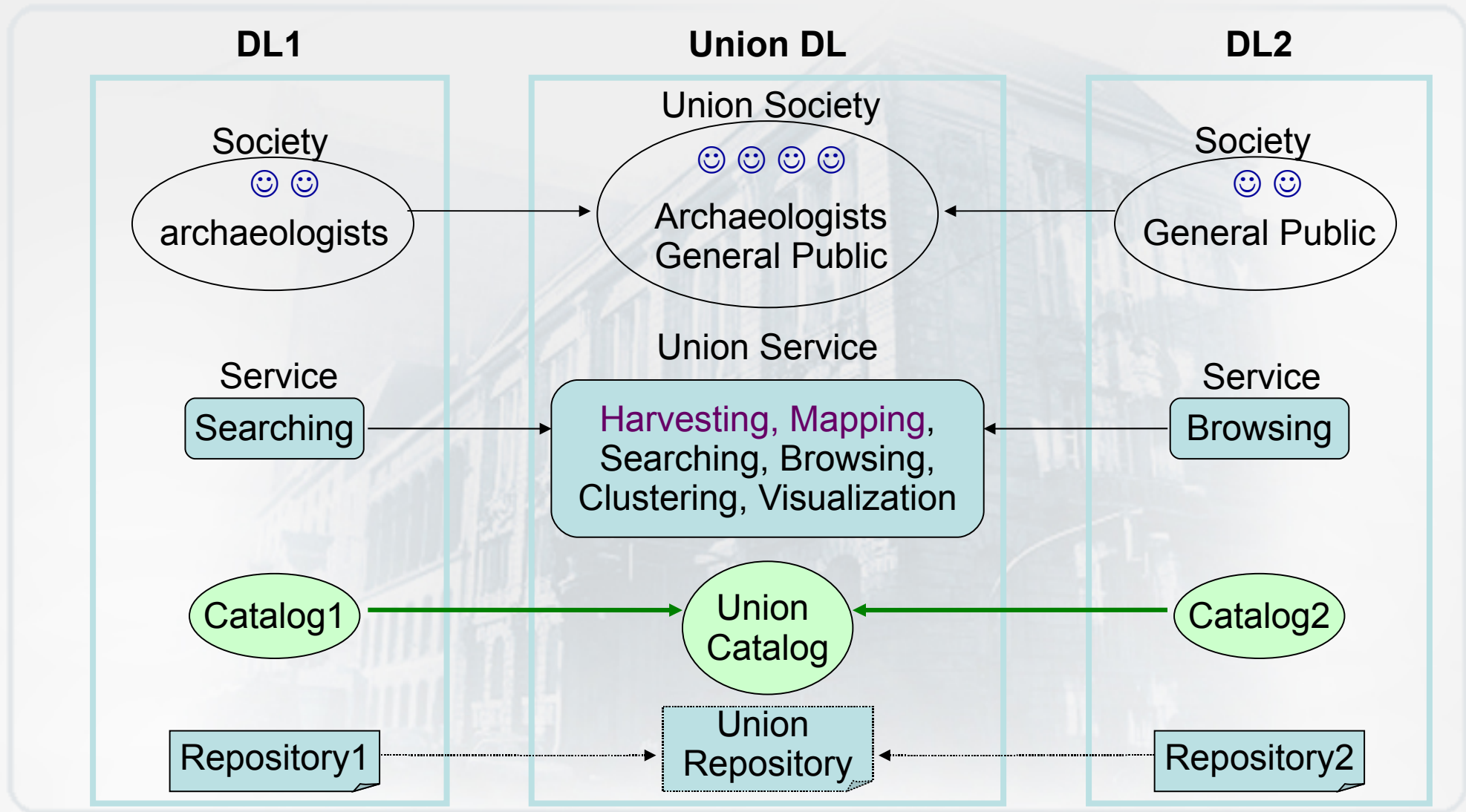
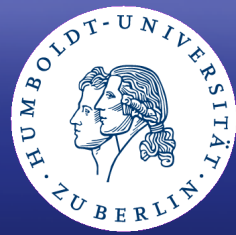


DL Integration: Frascati Proposal (Ed Fox)



- What is “DL Integration”
 - Hide distribution
 - Hide heterogeneity
 - Enable autonomy of individual component
- Why Integration
 - Island-DLs
 - inability to seamlessly and transparently access knowledge across DLs
 - Utilize various autonomous DLs in concert
 - e.g., toward The European DL !

A Union DL Architecture



Formal Definition of DL Integration

- $DL_i = (R_i, DM_i, Serv_i, Soc_i), 1 \leq i \leq n$
 - R_i is a network accessible repository
 - DM_i is a set of metadata catalogs for all collections
 - $Serv_i$ is a set of services
 - Soc_i is a society
- UnionRep
- UnionCat
- UnionServices
- UnionSociety
- DL integration problem definition: Given n individual libraries, integrate the n DLs to create a **UnionDL**.

5S Model: Characteristics



- Meta-Models and Ontologies may be useful components of a long term conceptual framework
- Potentially combined with DELOS framework components
- 5Ss themselves moderately useful ...
- 5S model is somewhat closer to Collective Memory Institutions' reality than DELOS
- Consider the Union DL vision!

DCMI Abstract

A Very Abstract Framework

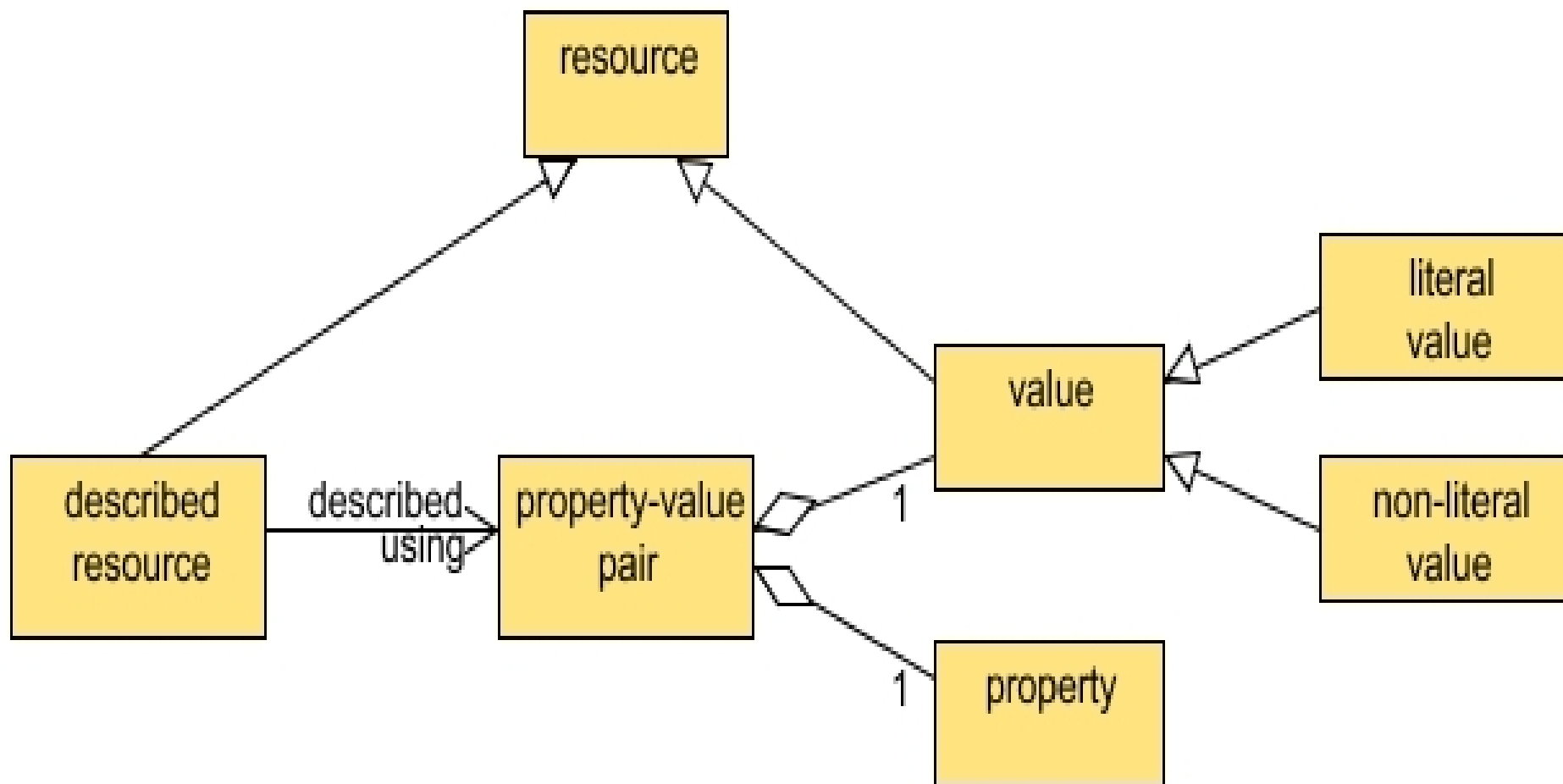
DCAM: Resources

- DCAM concerned with description of **resources**
- DCAM adopts Web Architecture/RFC3986 definition of resource
 - the term "resource" is used in a general sense for whatever might be identified by a URI. Familiar examples include an electronic document, an image, a source of information with consistent purpose (e.g., "today's weather report for Los Angeles"), a service (e.g., an HTTP to SMS gateway), a collection of other resources, and so on.
 - A resource is not necessarily accessible via the Internet; e.g., human beings, corporations, and bound books in a library can also be resources.
 - Likewise, abstract concepts can be resources, such as the operators and operands of a mathematical equation, the types of a relationship (e.g., "parent" or "employee"), or numeric values (e.g., zero, one, and infinity).

DCAM: Basics

- DCAM describes
 - **Components** and **constructs** that make up an information structure ("DC description set")
 - How that information structure is to be **interpreted**
- DCAM does not describe how to represent DC description set in concrete form
- DCAM describes various **types** of metadata terms, but does not specify the **use of any fixed set of terms**
- Made up of three related "information models"
 - **Resource** model
 - **Description** set model
 - **Vocabulary** model

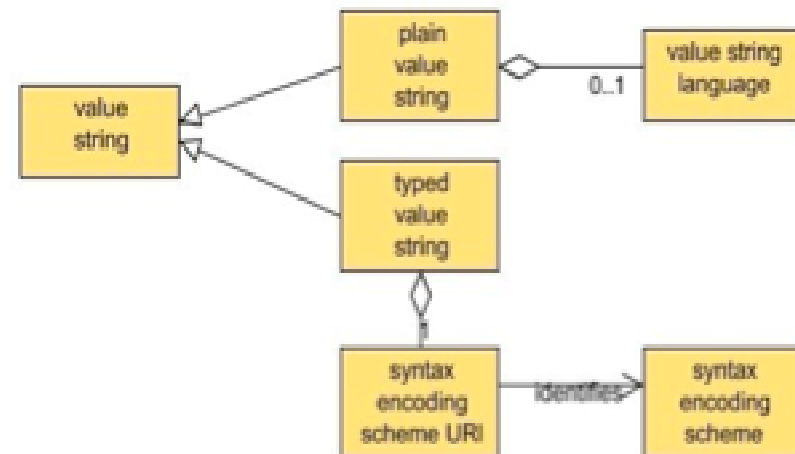
DCAM: Resource Model



DCAM: Resource Model

- The “view of the world” / conceptualisation on which DC metadata is based
- a described resource is described using one or more property-value pairs
- a property-value pair is made up of
 - exactly one property and
 - exactly one value
- a value is a resource
- a value is either a literal value or a non-literal value
- i.e. similar to RDF model of binary relations between resources; entity-relational model

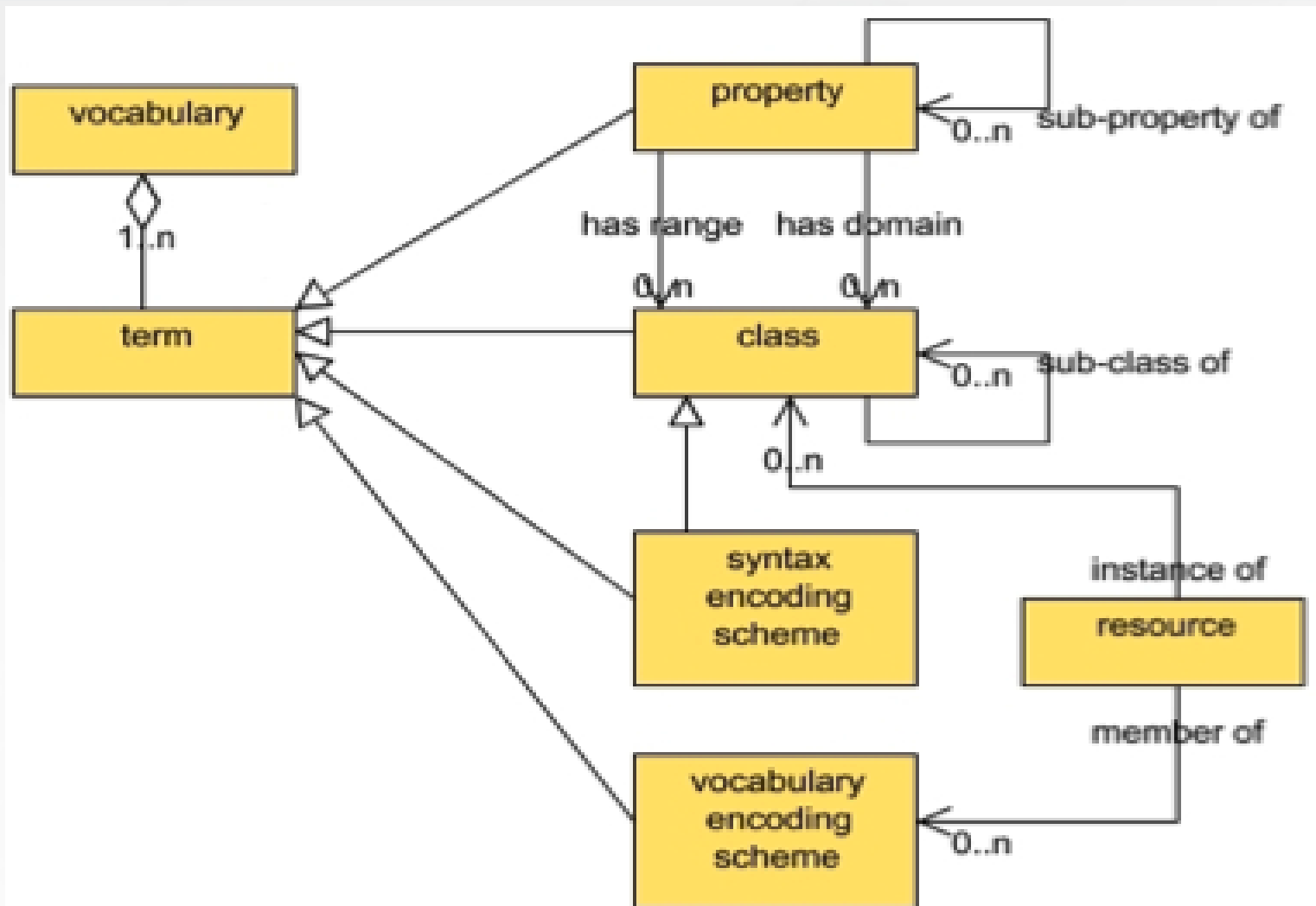
DCAM: Description Set Model (1)



DCAM: Description Set Model (2)

- The structure of DC metadata
- Uses URIs to refer to resources & metadata terms (like RDF)
- a description set is made up of one or more descriptions, each of which describes one resource
- a description is made up of
 - zero or one described resource URI
 - identifies described resource
 - one or more statements
- a statement is made up of
 - exactly one property URI
 - identifies property
 - exactly one value surrogate
- a value surrogate is either a literal value surrogate or a non-literal value surrogate

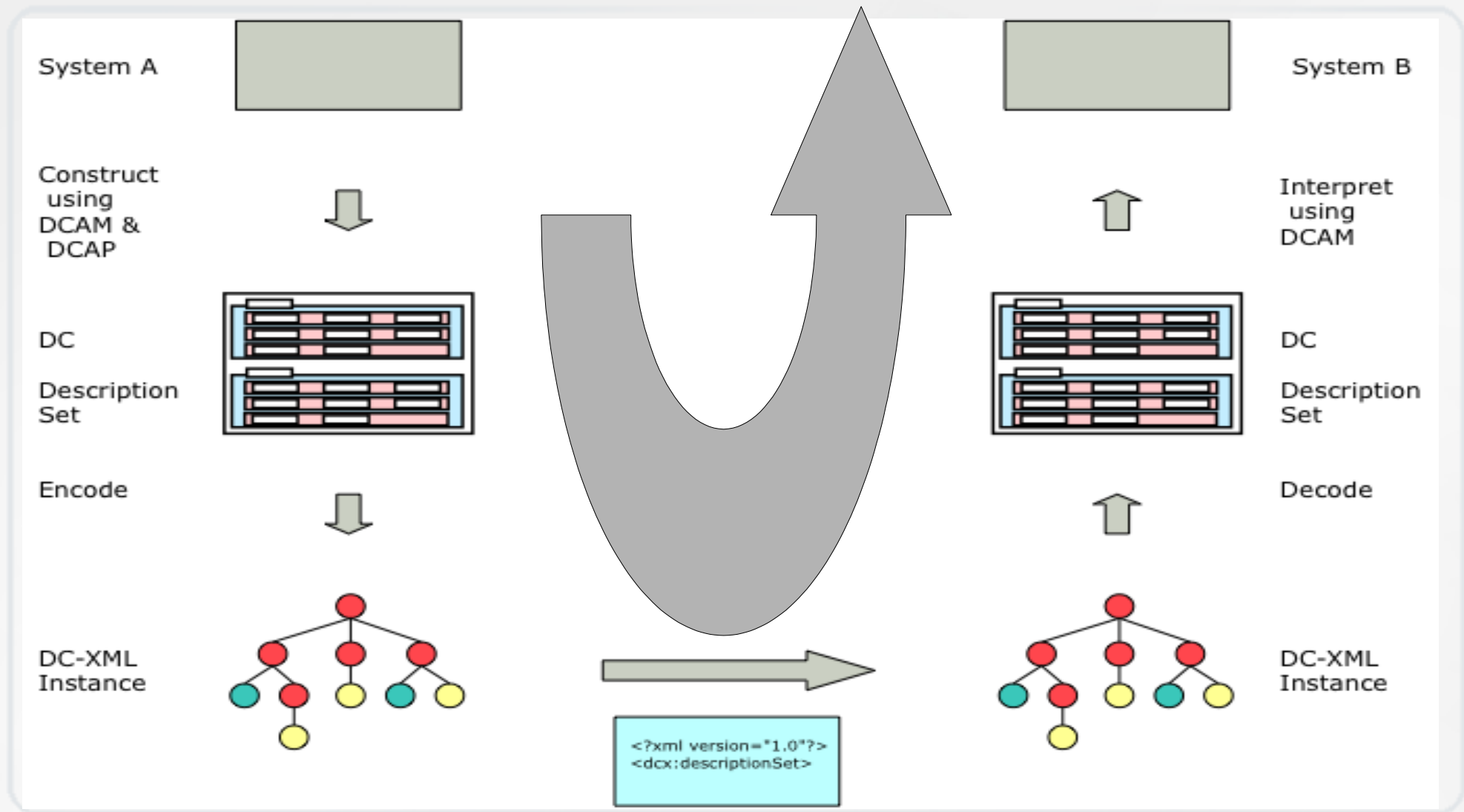
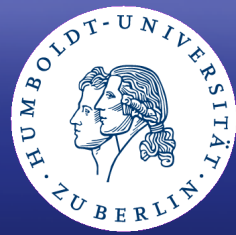
DCAM: Vocabulary (1)



DCAM: Vocabulary (2)

- a **vocabulary** *is* a set of terms (property, class, vocabulary encoding scheme, syntax encoding scheme)
- a **resource** *may be* an instance of one or more classes
- a **resource** *may be* a member of one or more vocabulary encoding schemes
- a **property** *may have* a range relationship with one or more classes
- a **property** *may have* a domain relationship with one or more classes
- a **property** *may have* a subproperty relationship with one or more properties
- a **class** *may have* a subclass relationship with one or more classes
- \sim RDF Schema

DCMI Abstract and Interoperability



DCMI Abstract: Characteristics

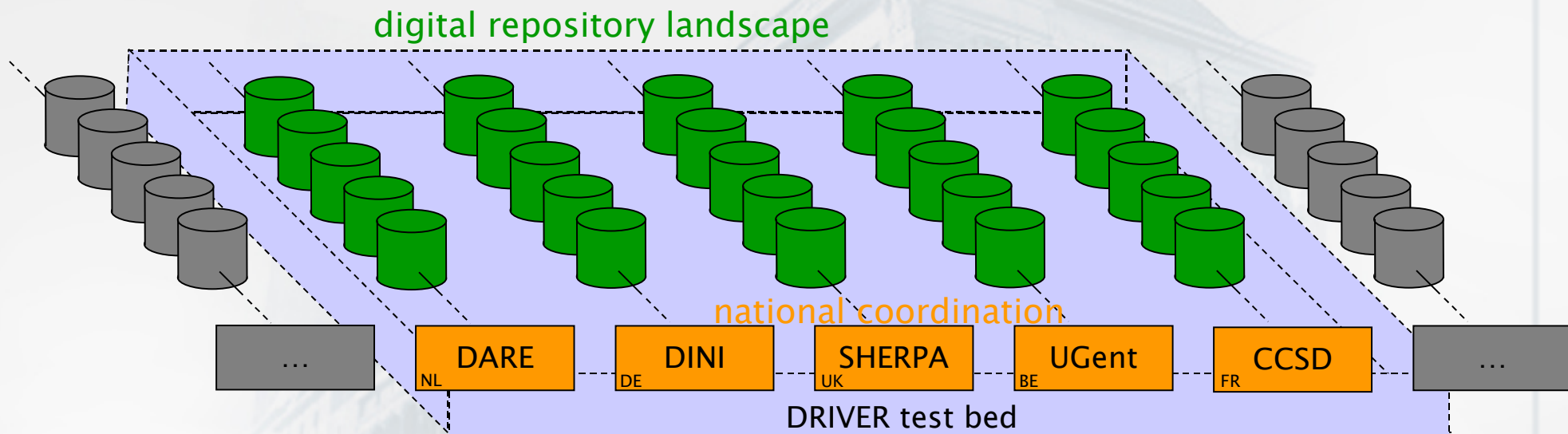
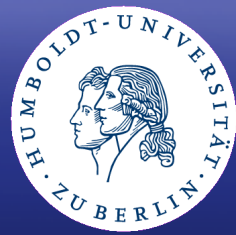


- Conceptually useful approach
- Relatively weak acceptance
- Probably too abstract to be considered in operational settings
- Combine with DC:Terms to create strong metadata interoperation framework
- Break!

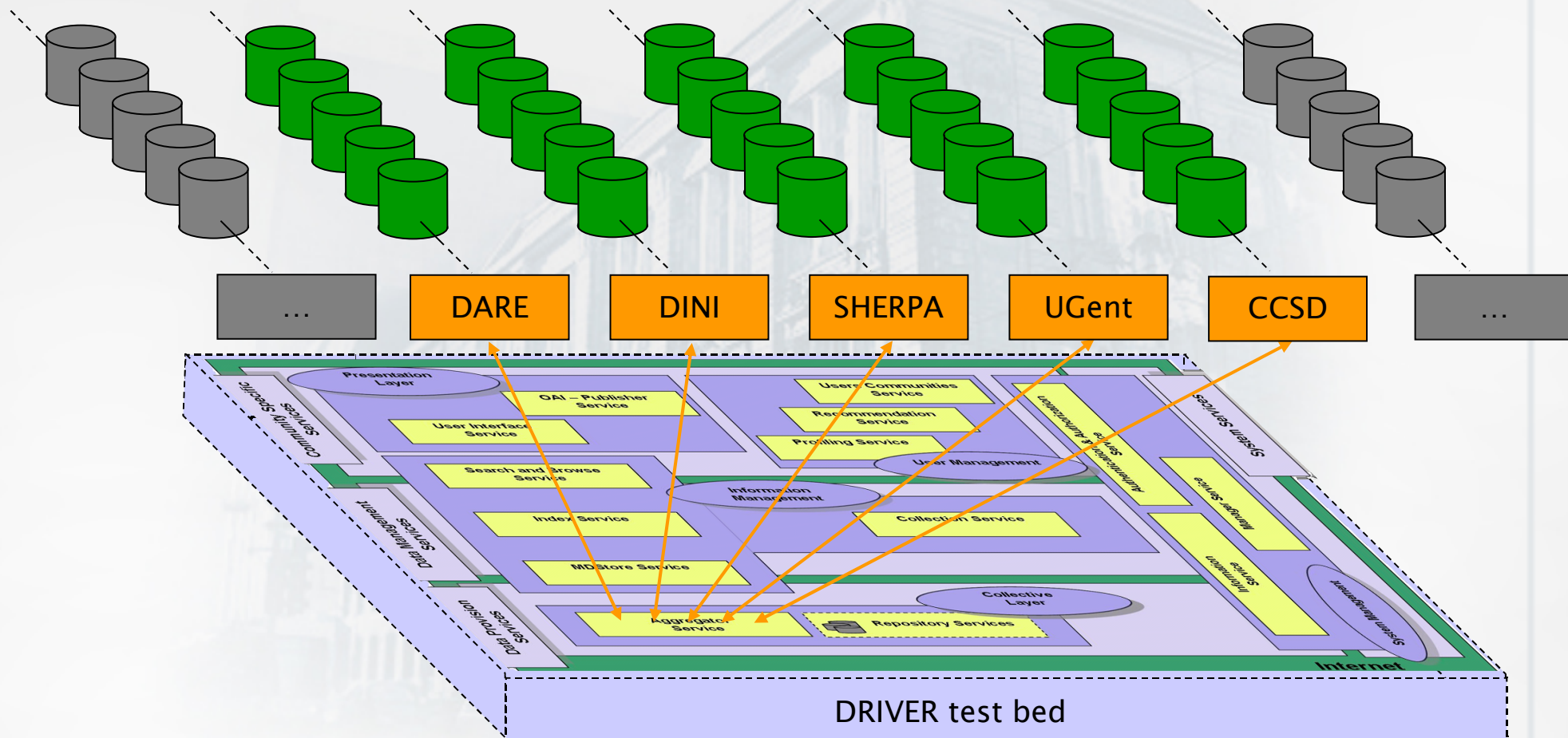
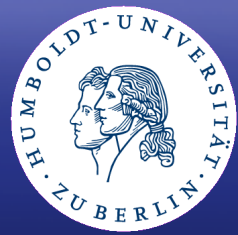
DRIVER

Federated Repositories Harvesting⁺⁺

DRIVER Test Bed Organisational Structure

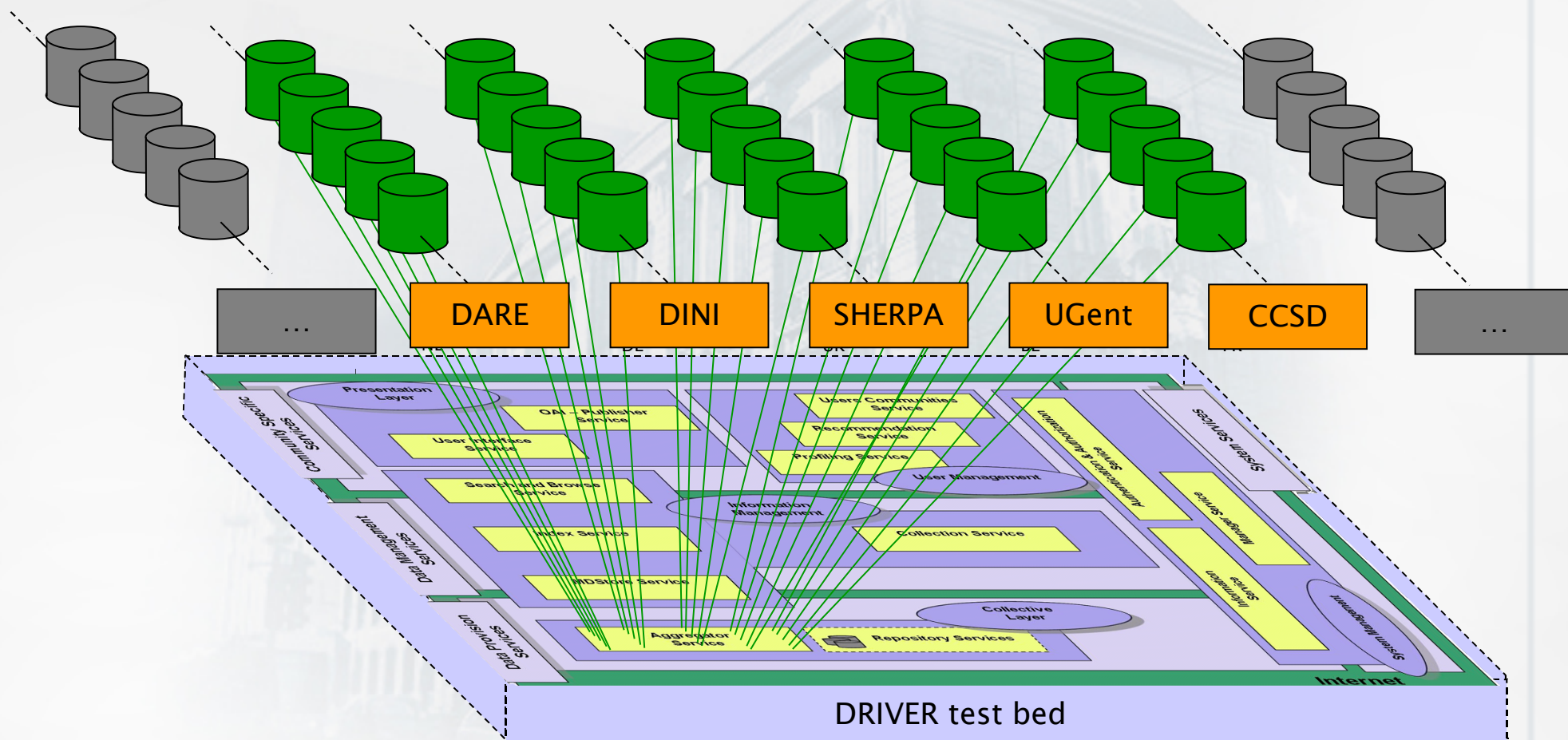


Harvesting Strategy #1: National Aggregators



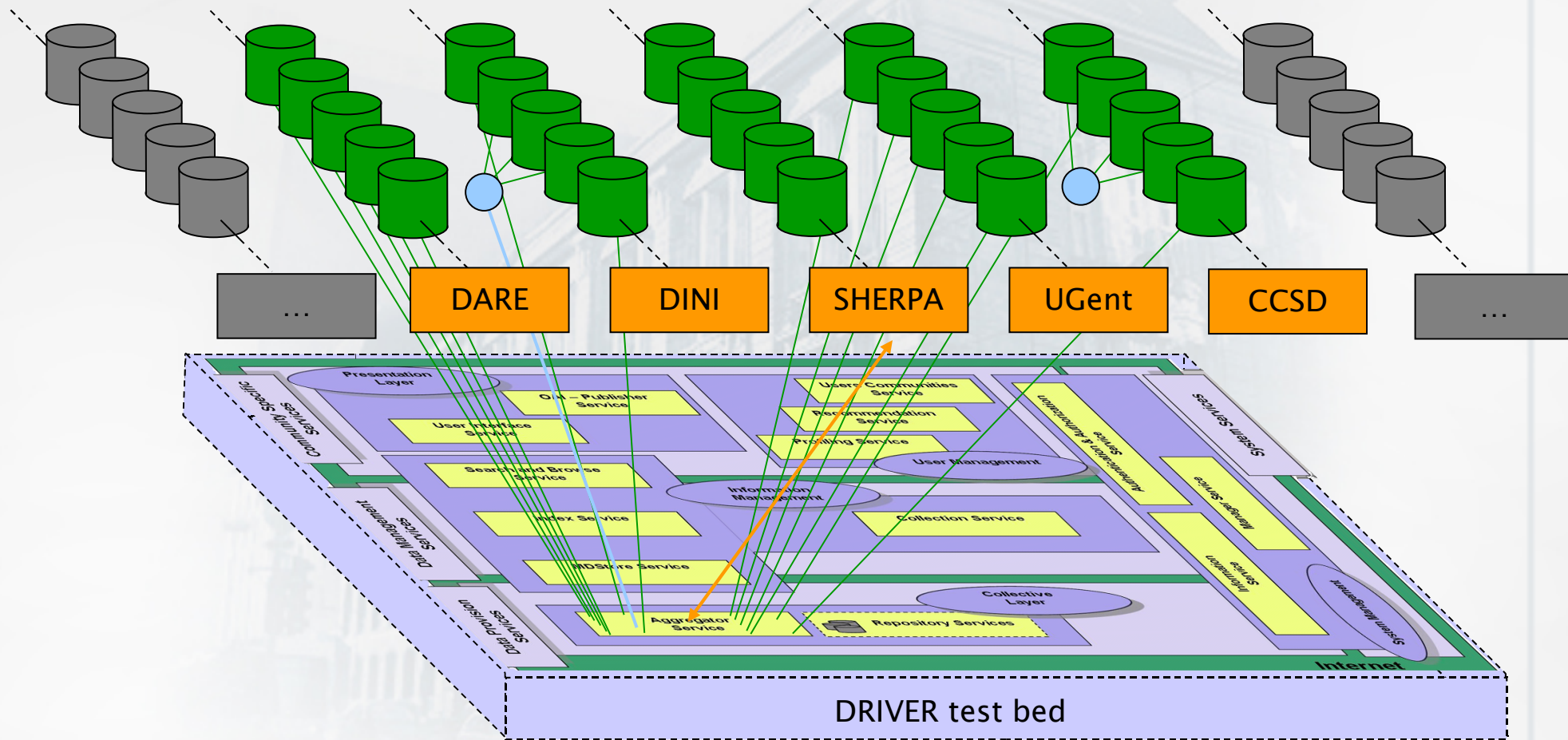
DL-Interoperability@Firenze 25-09-2008

Harvesting Strategy #2: Local Harvesting



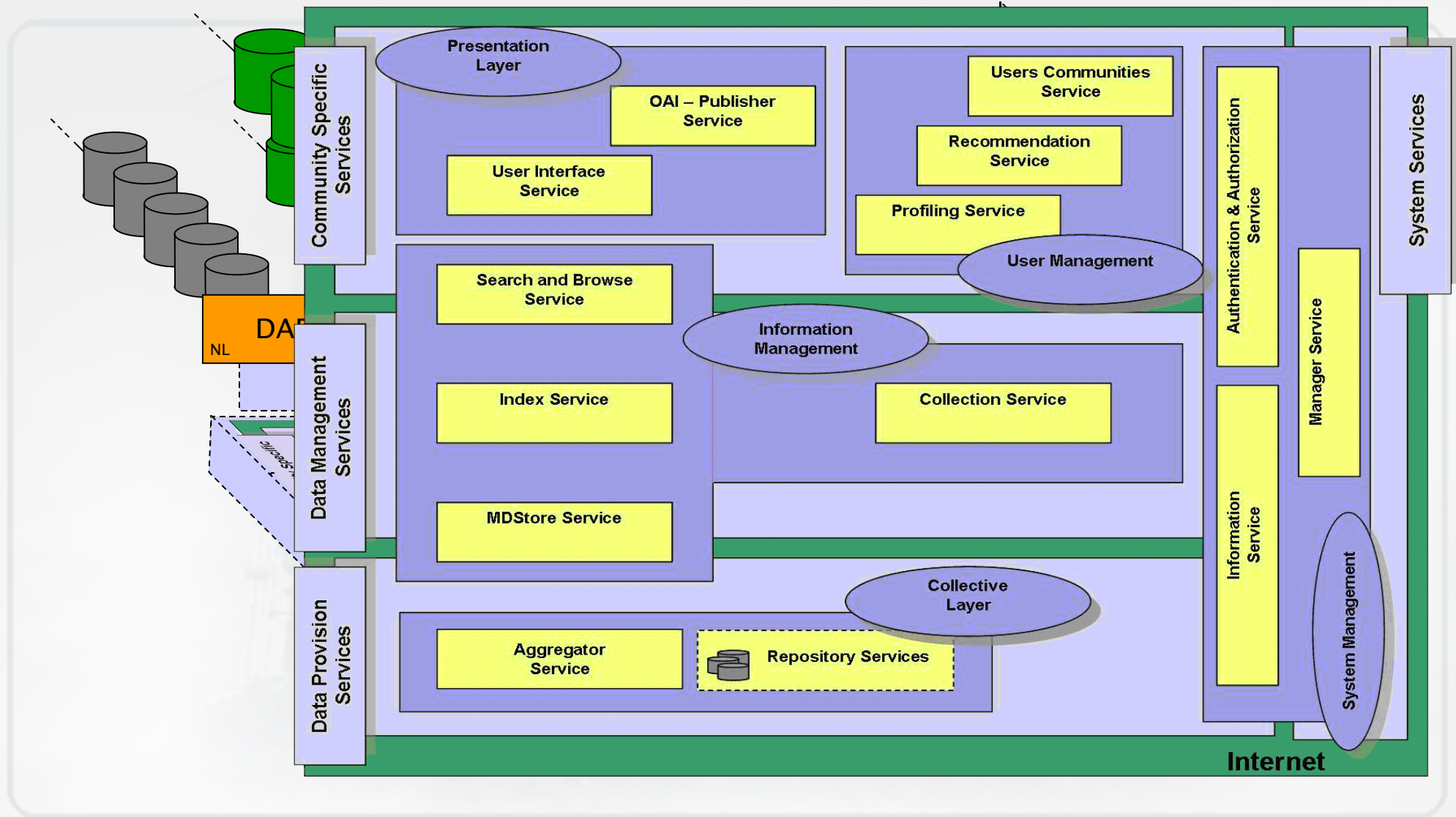
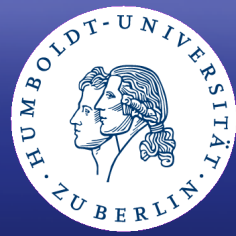
DL-Interoperability@Firenze 25-09-2008

Harvesting Strategy #3: Eclectic Harvesting



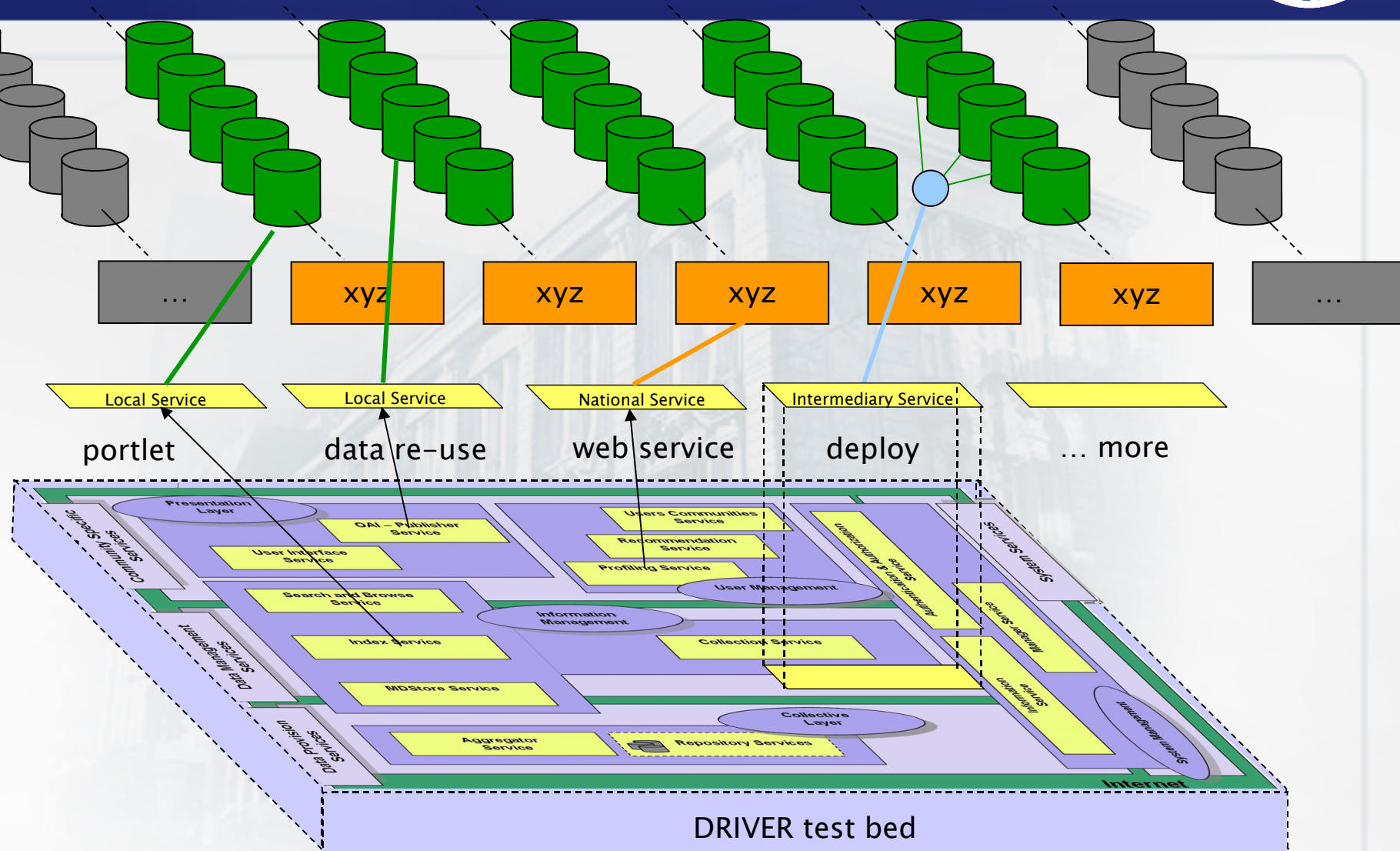
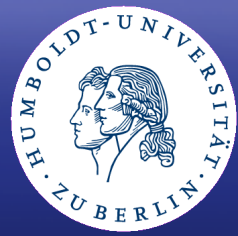
DL-Interoperability@Firenze 25-09-2008

DRIVER Open Service Architecture



DL-Interoperability@Firenze 25-09-2008

DRIVER Service Interaction



DL-Interoperability@Firenze 25-09-2008

DRIVER: Characteristics

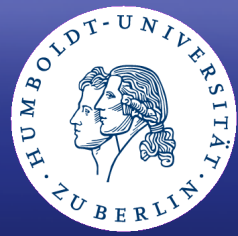


- OAI Harvesting+
- Value added services on top of aggregated repository content
- Harvesting based model of repository federation
- Limited set of core functions
- Limited to textual objects, but currently being extended to complex and multimedia objects in DRIVER2
- => Infrastructure framework for providing platform interoperability

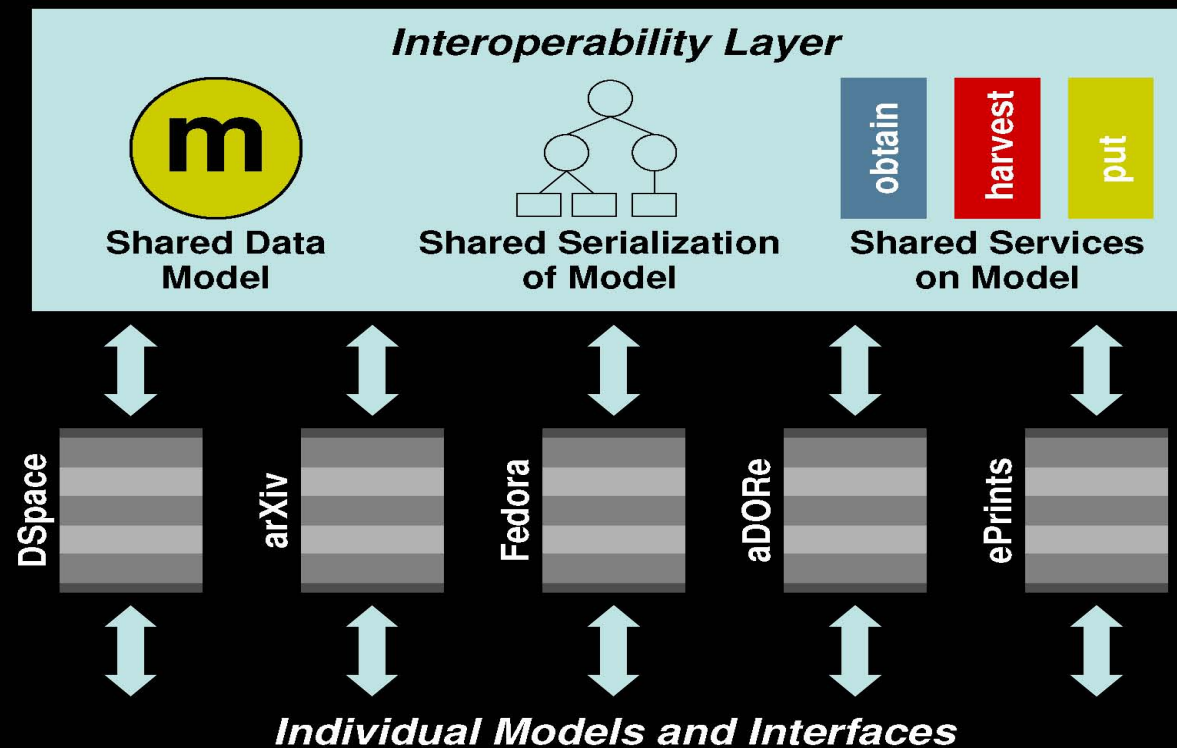
OAI-ORE

Generic Internet/WWW Methodology Providing Interoperability

Pathways / OAI-ORE: Data Model & 3 Functional Primitives



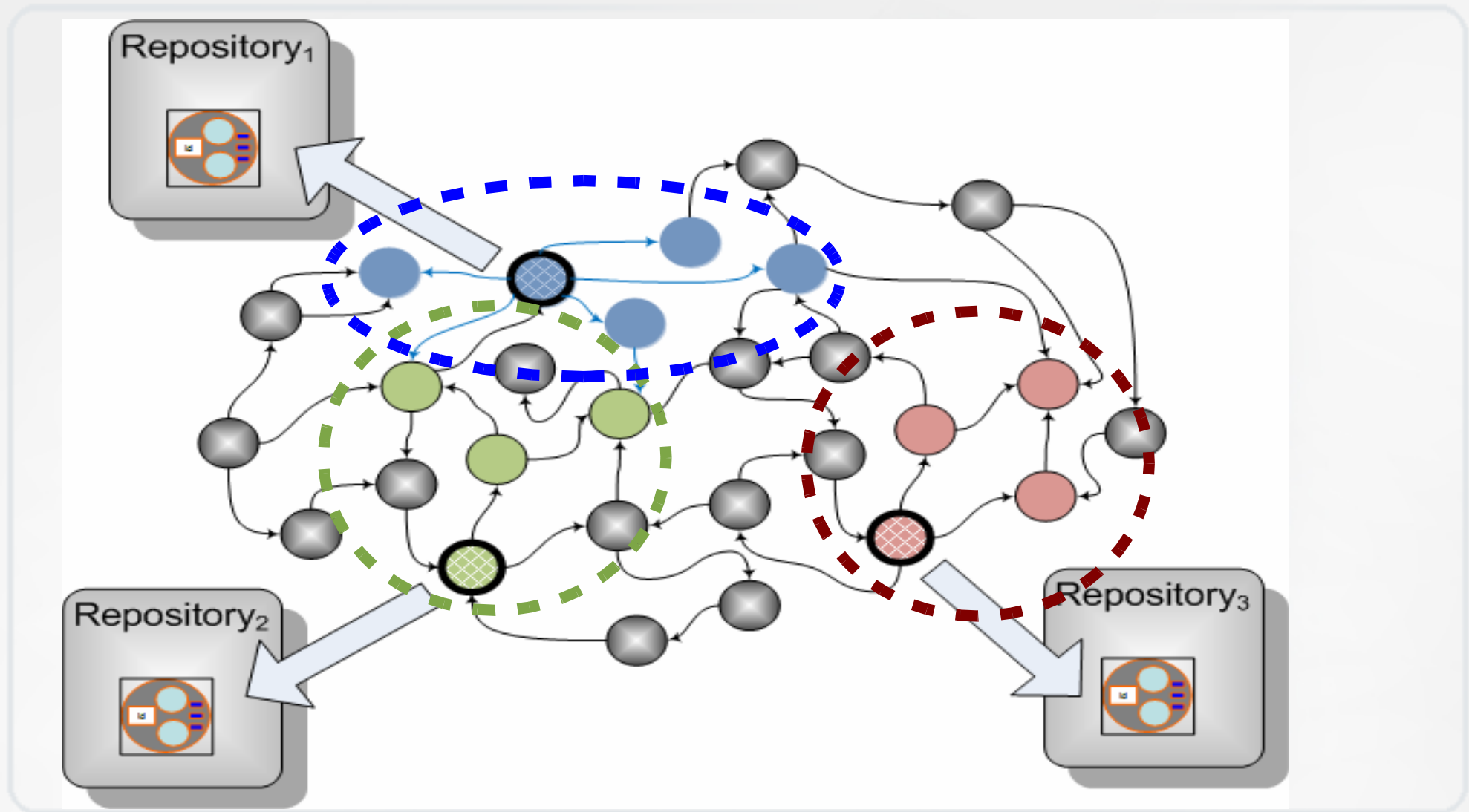
The Pathways interoperability fabric



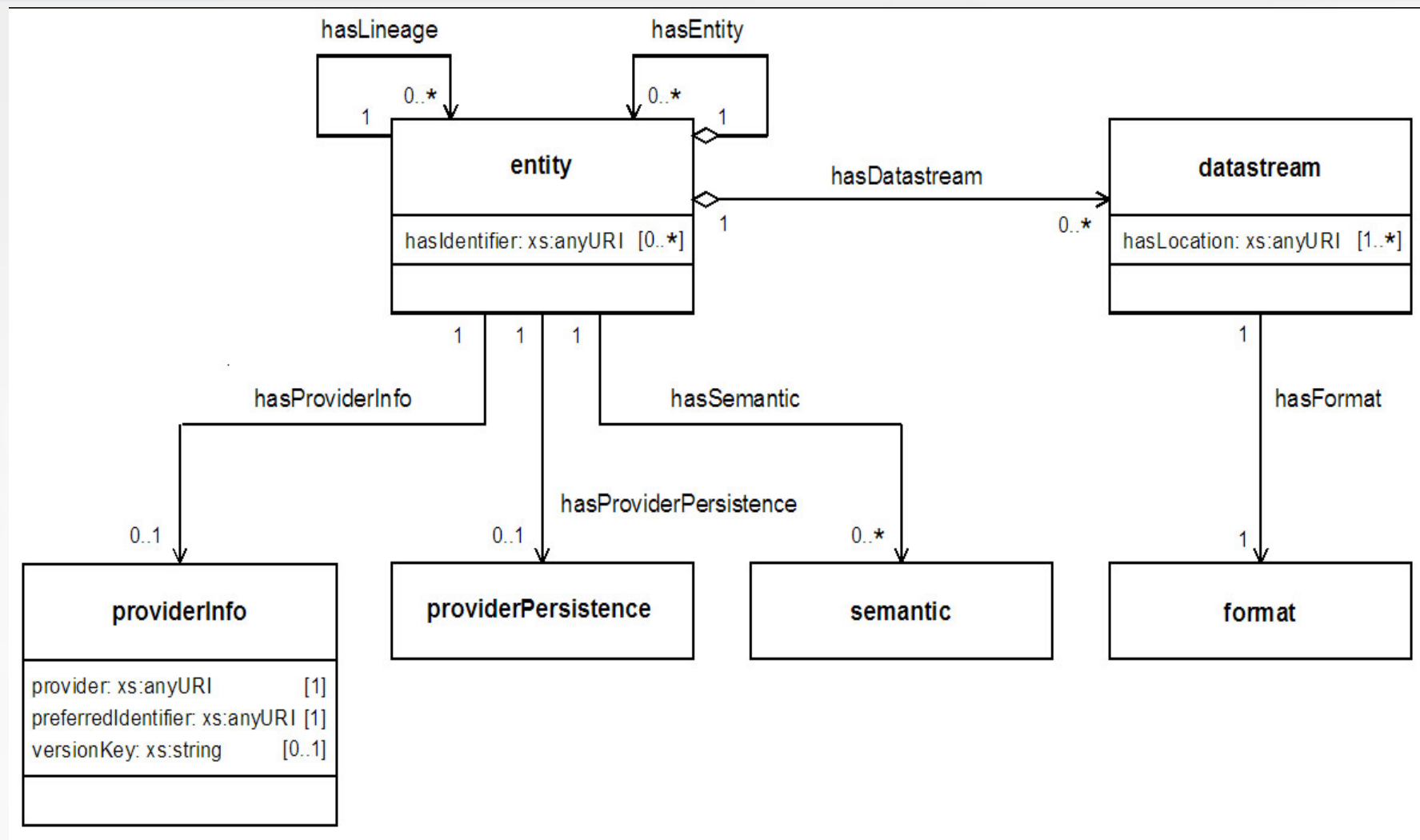
OAI ORE: Principles

- Goal: „Facilitate Use and Re-Use of Compound Information Objects (and of their component parts)“
- „How to deal with compound information objects in a manner that is in sync with the Web architecture?“
- By enriching the web graph with boundary information.

... web graph with boundary information



Pathways / OAI-ORE: Data Model (initial!)

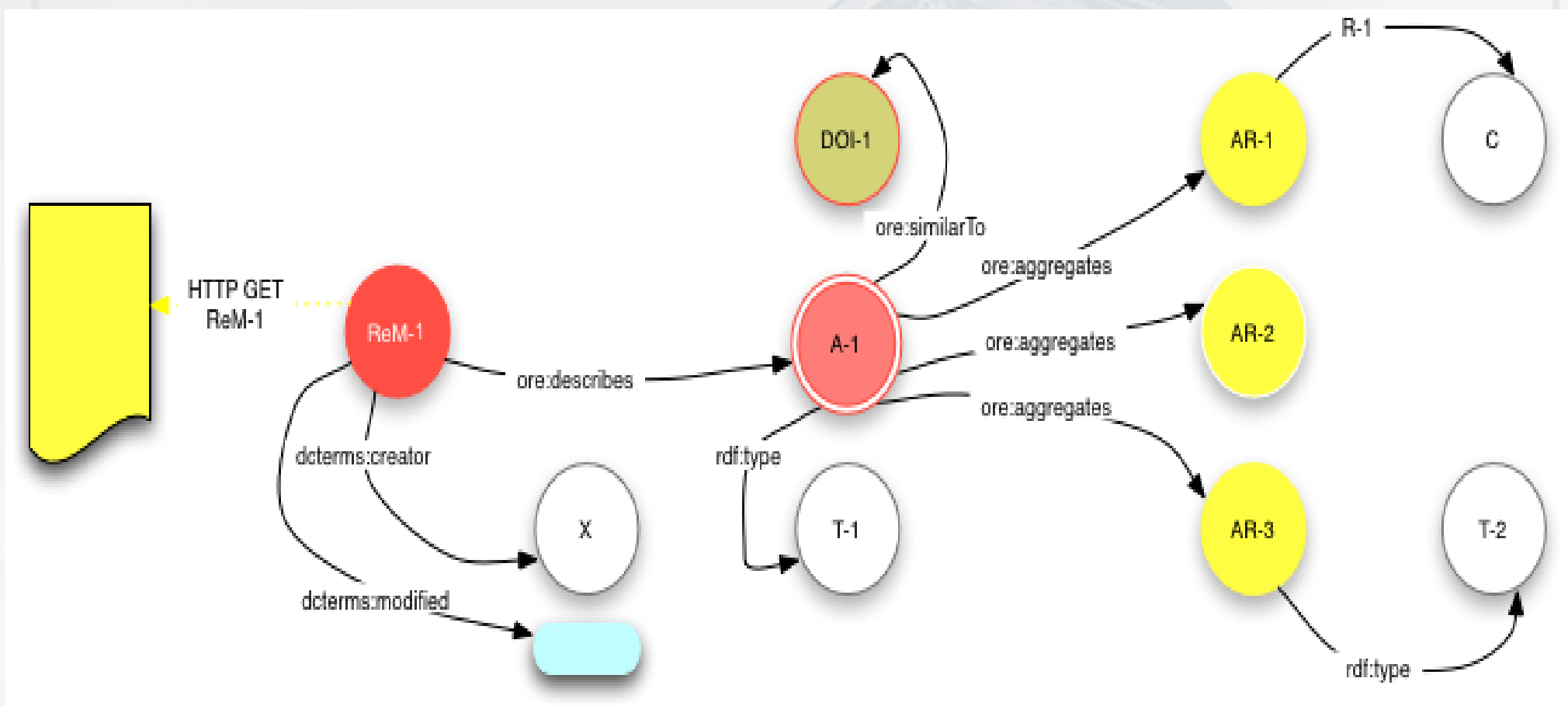


ORE as an Interoperability layer for compound information objects



- The anticipated interoperability layer for compound information objects consists of approaches to facilitate:
 - The **publication of named graphs** to the web as a means to convey compound object (i.e. boundary) information.
 - **Discovery** of these named graphs.
 - **Assessment of the trustworthiness** of named graphs as an information source.
 - Development of a variety of **vocabularies for expressing types of links** between resources denoted by the nodes in a named graph.
 - Development of a variety of **vocabularies for expressing properties** of resources denoted by nodes in a named graph, especially semantic type, media type, and media format.
- Bootstrap vocabularies only for italics sections

An Example Resource Map



ORE: Resource Maps

- A Resource Map is the **serialization of a named graph** that corresponds with a compound object.
- Resource Maps must allow for simply expressing the resources that are considered **part of** a compound object.
- Resource Maps may
 - Express resources that are **not part** of a compound object.
 - Distinguish between resources that **are part** of the compound object and those that **are not**.
 - Express the **relationships among the resources** referenced by the named graph.
 - Express the **types of the relationships** among the resources referenced by the named graph, i.e. label the arcs.
 - Express other **information related to the named graph** and to the resources that it references such as metadata, etc.

ORE: Characteristics



- Limited to exchange of objects!
- But very useful for this limited
- 100% based on standard, generic WWW technology
- Does not address/answer some fundamental issues such as boundary constitution
- => Relatively sure bet for persistent web based object modelling

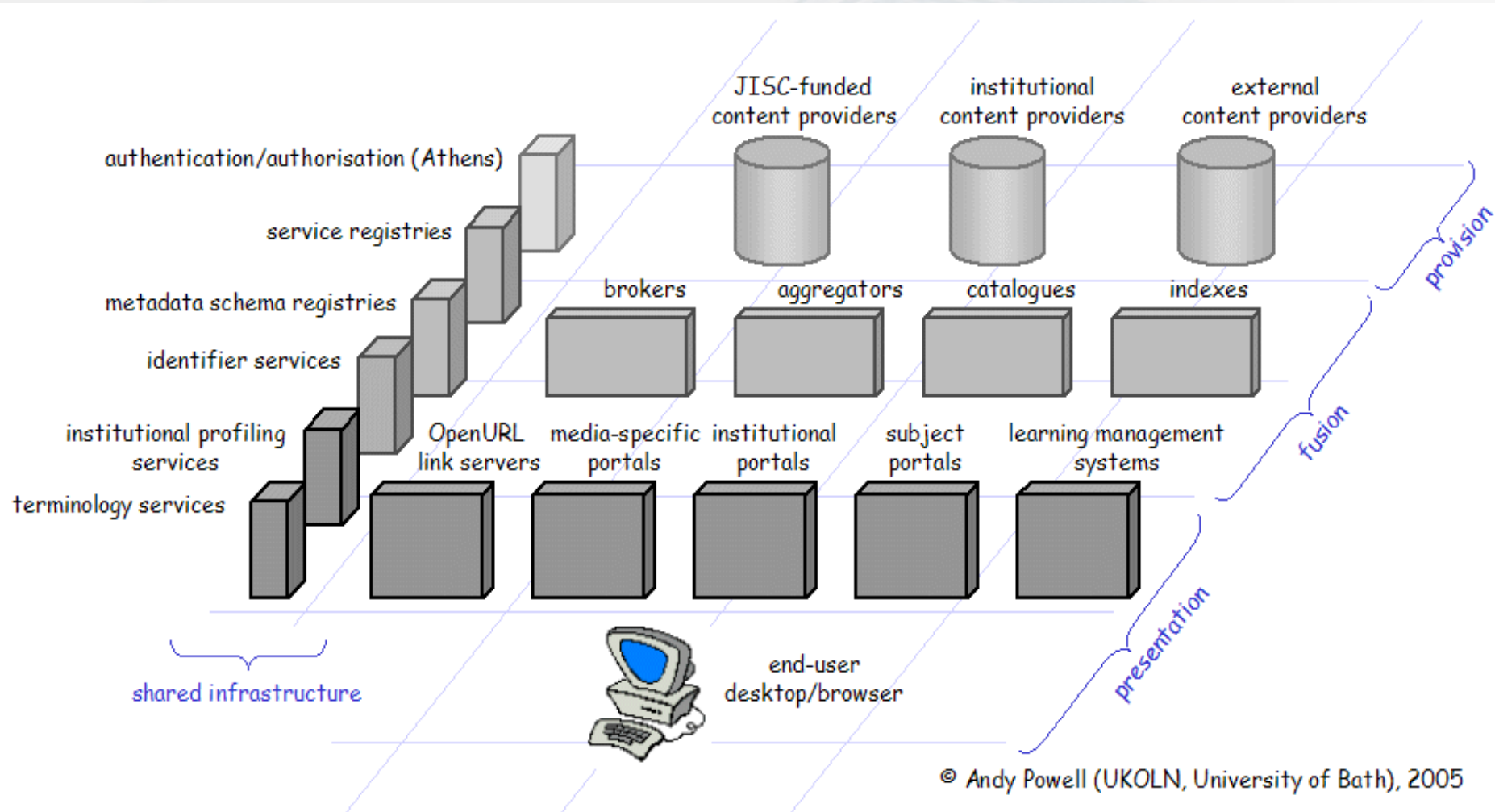
JISC Information Environment

Service Oriented Architecture for Linking UK Repositories

JISC Information Environment



- <http://www.ukoln.ac.uk/distributed-systems/jisc-ie/arch>



DL-Interoperability@Firenze 25-09-2008

JISC IE: Service Model for Linking Repositories

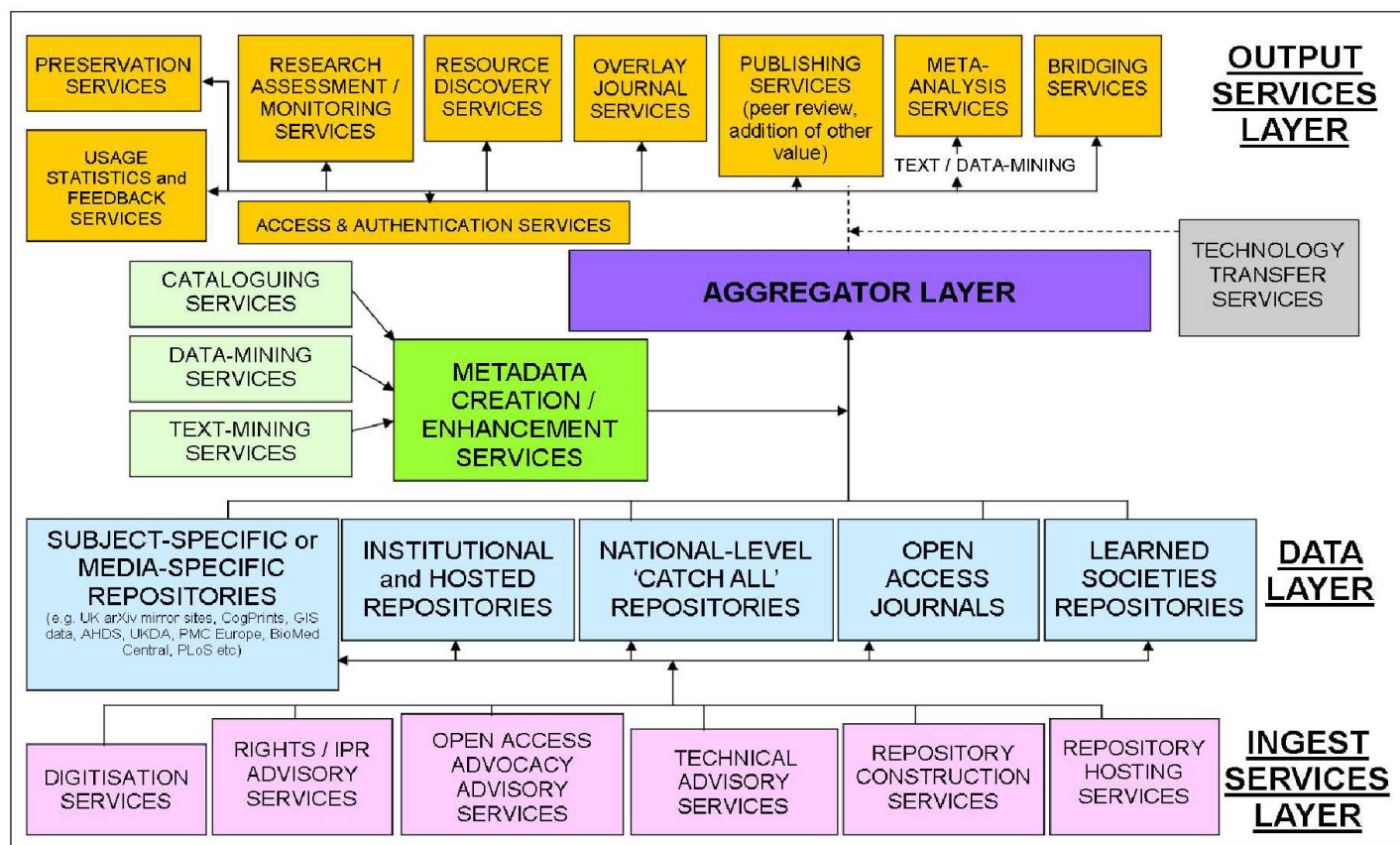


Chart A: Overall model for repositories and the services built across them

JISC IE: Characteristics



- Exclusively SOA oriented
- Objective is to „ ... support user-oriented services across digital repositories ...“ (Swan 2006)
- Service model is quite close to 'librarian' reality, even though explicitly designed for repositories

JCR (JSR170/283)

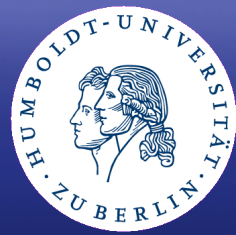
Content Infrastructure with High Functional Granularity

JCR: Functionality Overview



- **Granular Read/Write Access** - This is the bi-directional interaction of content elements. Issues with access on a property level and not just on a "document" level
- **Versioning** - Transparent version handling across the entire content repository, providing the ability to create versions of any content and select versions for any content access or modification.
- **Hard- and Soft-structured Content** - An Object Model that defines how hard and soft-structured content could be addressed.
- **Event Monitoring (Observation)** - Possible use of JMS based notification framework allowing for subscription on content modification.
- **Full-text Search and filtering** - Entire (non-binary) repository content indexed by a full-text search engine that enables exact and sub-string searching of content.
- **Access Control** - Unified, extensible, access control mechanisms.
- **Namespaces & Standard Properties** - Defining default standard properties that will maintain namespace uniqueness and hierarchy.
- **Locking and Concurrency** - Standardized access to locking and concurrency features
- **Linking** - A standard mechanism to soft/hard link items and properties in a repository along with providing a mechanism to create relationships in the repository.
- ... more: [Specification Document](#)

JCR: the Community



- OASIS Expert Group Membership:
 - **DMS/CMS industry:** BEA Systems, Day Software, Inc., Documentum, Inc., Filenet Corporation, Hummingbird Ltd., Interwoven, Opentext, RedDot Solutions AG, Stellent, Inc., Venetica Corporation, Vignette
 - **Major Technology Providers:** Apache Software Foundation, Art Technology Group Inc.(ATG), Fujitsu Limited, Hewlett-Packard, IBM, Macromedia, Inc., Mediasurface Ltd., Novell, Inc., Oracle, SAP AG, SAS Institute Inc., Software AG, Sun Microsystems, Inc.,

JCR: Characteristics

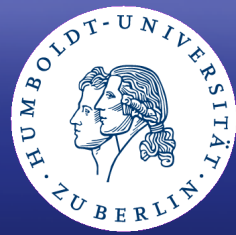


- “ ... lay the foundations for a true industry-wide content infrastructure”: focus is on interaction with DMSs
- Industry standard: a very different type of community!
- Limited take-up in industry
- Immediate impact may be limited because of limitation to Java implementation ...
- ... but offers a very granular set of functional primitives!

iRODS

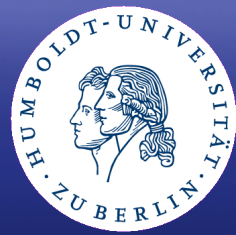
Data grid system for interoperability

iRODS/SRB: Big Picture



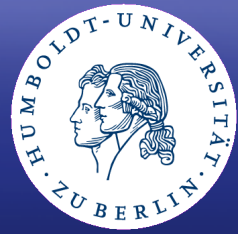
- iRODS stands for **integrated Rule-Oriented Data Systems**.
- Second generation **data grid system** providing a unified view and seamless access to **distributed digital objects** across a wide area network.
- Storage Resource Broker (SRB):
 - first generation data grid system providing a **unified view based on logical naming concepts** - users, resources, data objects and virtual directories were abstracted by logical names and mapped onto physical entities -
 - providing a physical-to-logical independence for client-level applications.
- iRODS builds upon this logical abstraction and takes it one level higher by abstracting the data management process itself called **policy abstraction**.

iRODS: the 'Integrated Envelope' Paradigm



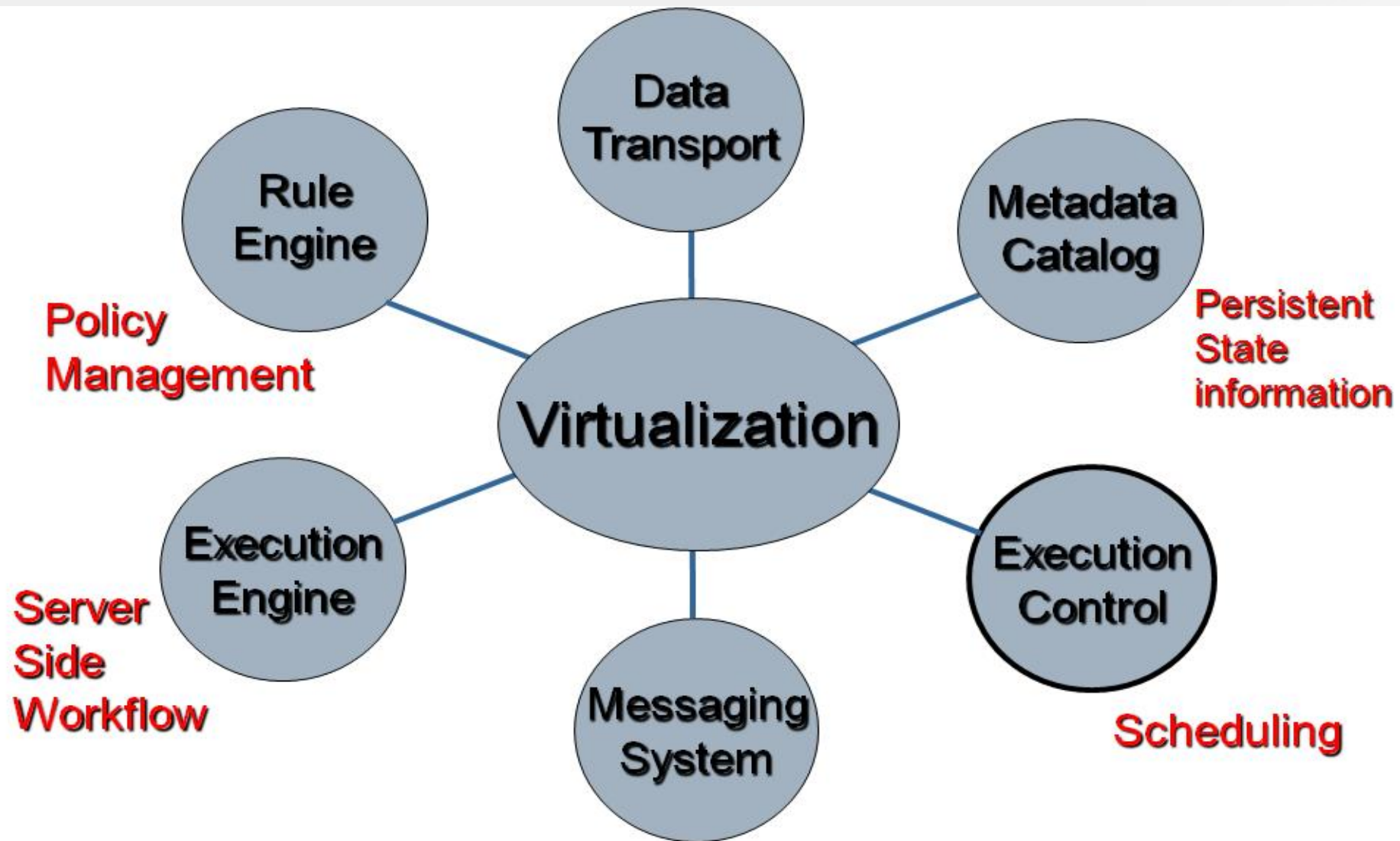
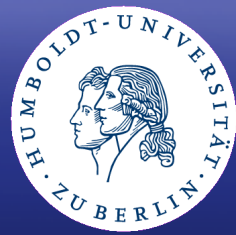
- 'Integrated Envelope' provides a **unified environment for interactions with a host of underlying services** which interact in complex fashion among themselves.
- Exposes a **uniform interface to the client application** hiding the complexity of dealing with low-lying details of the 'tools' inside the envelope.
- SRB integrated envelope: single-server installation paradigm **hiding** the details about third-party authentication, authorization, auditing, metadata management, streaming access mechanism, resource (vendor-level) and other idiosyncrasies.
- SRB offers around 100 API functions and 80 command-level utilities.
- iRODS builds upon the integrated envelope paradigm but with more **functionality** and **services**.

iRODS: Functionalities



- iRODS integrates the following functionalities:
 - **Data Transport**
 - **Metadata Catalog** for both system and user-defined metadata
 - **Rule Engine** for executing complex policies encoded as micro-services
 - **Execution Engine** for execution of remote micro-services as workflows
 - **Scheduling System** for immediate, delayed and periodic queuing and execution
 - **Messaging System** for out-of-band communication among micro-services
 - **Virtualization system** enabling the logical naming paradigm

iRODS: Functionalities



iRODS: Microservices



- Micro services are small, well-defined **procedures/functions that perform a certain task**
- Users and administrators can chain these micro-services to implement a larger **macro-level functionality** that they want to use or provide for others.
- The task that is performed by a micro-service can be quite small or very involved. We leave it to the micro-service developer to choose the proper **level of granularity** for their task differentiation.
- Sample micro-services are “createCollection”, “assignAccess”, “createPhysicalFile”, “computeChecksum” and “replicateObject”.

iRODS: More Microservices



- **Workflow Services:**
 - nop, null - no action
 - cut - not to retry any other applicable rules for this action
 - succeed - succeed immediately
 - fail - fail immediately - recovery and retries are possible
- **System Micro Services - Can only be called by the server process.**
 - msiSetDefaultResc - set the default resource
 - msiSetRescSortScheme - set the scheme for selecting the best resource to use
 - msiSetDataObjPreferredResc - specify the preferred copy to use in case of multiple copies
 - msiSetDataObjAvoidResc - specify the copy to avoid
- **User Micro Services - can be called by client through irule.**
 - msiDataObjCreate - create a data object
 - msiDataObjOpen - open a data object
 - msiDataObjRead - read an opened data object
 - msiDataObjWrite - write
 - msiDataObjUnlink - delete
 - msiDataObjCopy - copy
 - msiDataObjRename - rename a data object

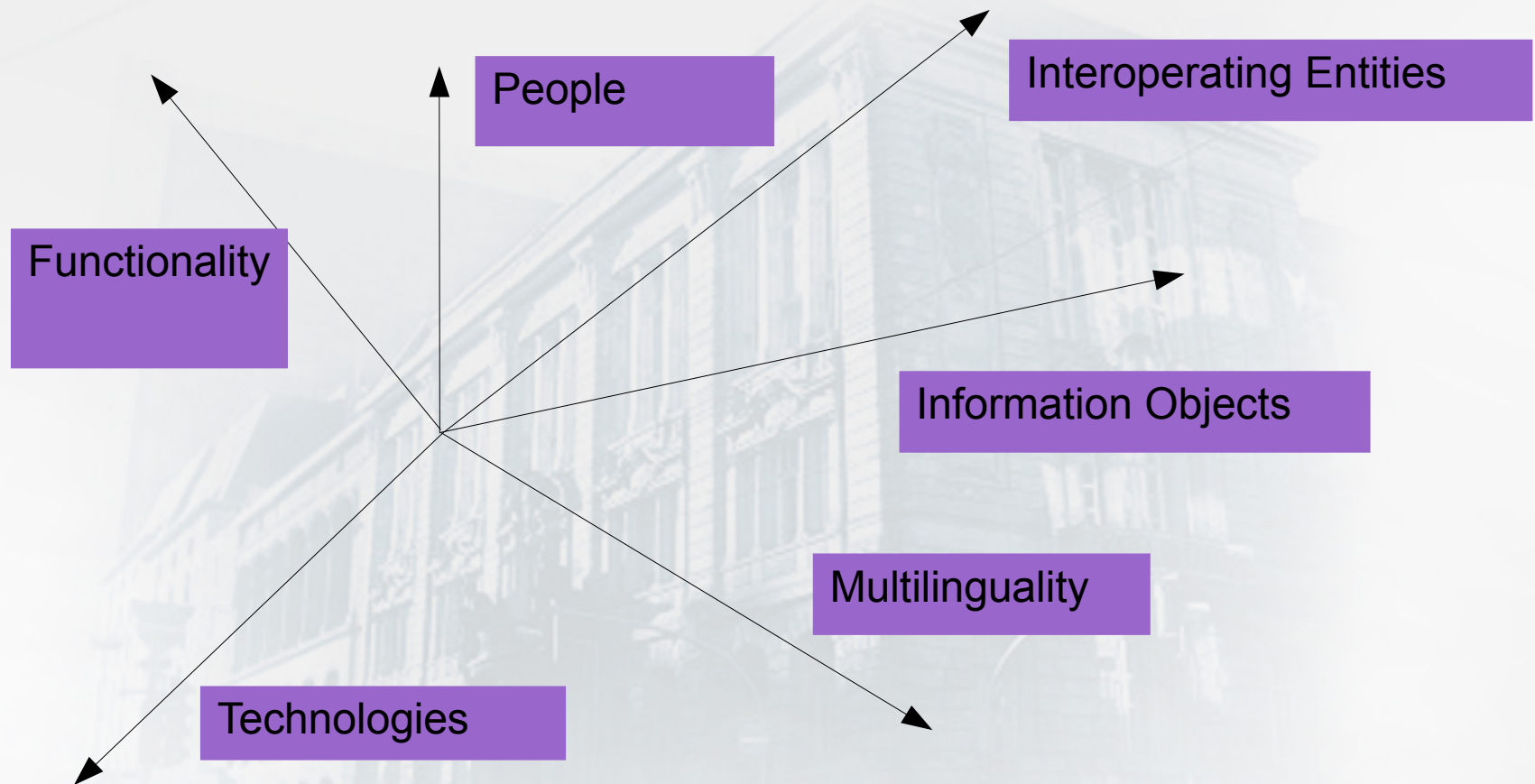
- Rules are **definitions of actions** (or macro-level tasks) that need to be performed by the server.
- These definitions are made in terms of micro-services and other actions.
- Basically a rule is specified with a line of text which contains 4 parts separated by the '|' separator:
actionDef | condition | workflow-chain | recovery-chain
 - **'actionDef'** is the name of the rule. It is an identifier which can be used by other rules or external functions to invoke the rule.
 - **'condition'** is the condition under which this rule applies. i.e., this rule will apply only if the condition is satisfied.
 - **'workflow-chain'** is a sequence of micro-services/rules to be executed by this rule.
 - **'recovery-chain'** are the rules to be called when execution of any one of the rules in the workflow-chain failed.

iRODS: Characteristics



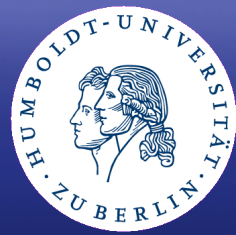
- Server based platform for application development
- Hiding distribution
- Centralised, unified API (unlike web services!)
- Relatively low abstraction level
- Not well suited for legacy systems

Interoperability: 6 vectors on 4 abstraction levels



DL-Interoperability@Firenze 25-09-2008

Interoperability Abstraction Levels (Tolk 2006)



Composability

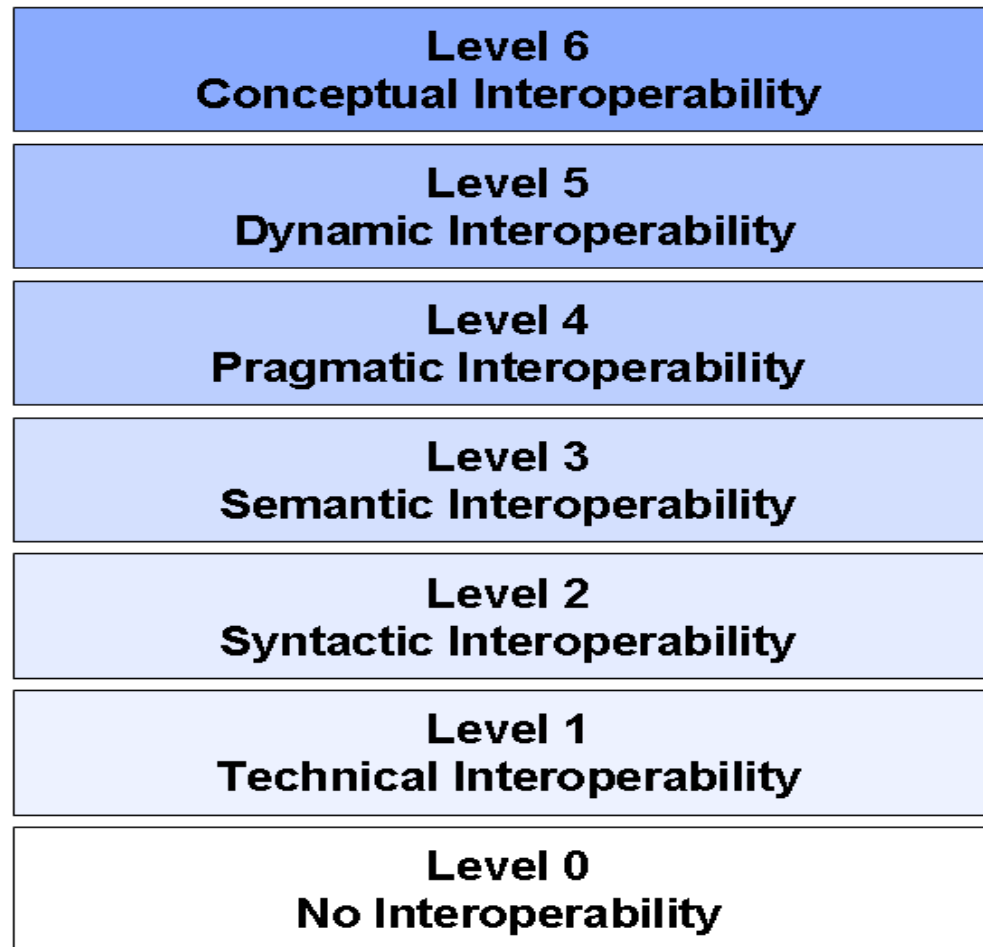
**Modeling /
Abstraction**

Interoperability

**Simulation /
Implementation**

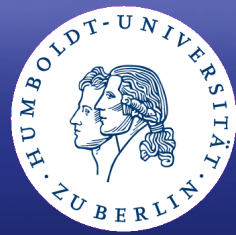
Integrability

**Network /
Connectivity**

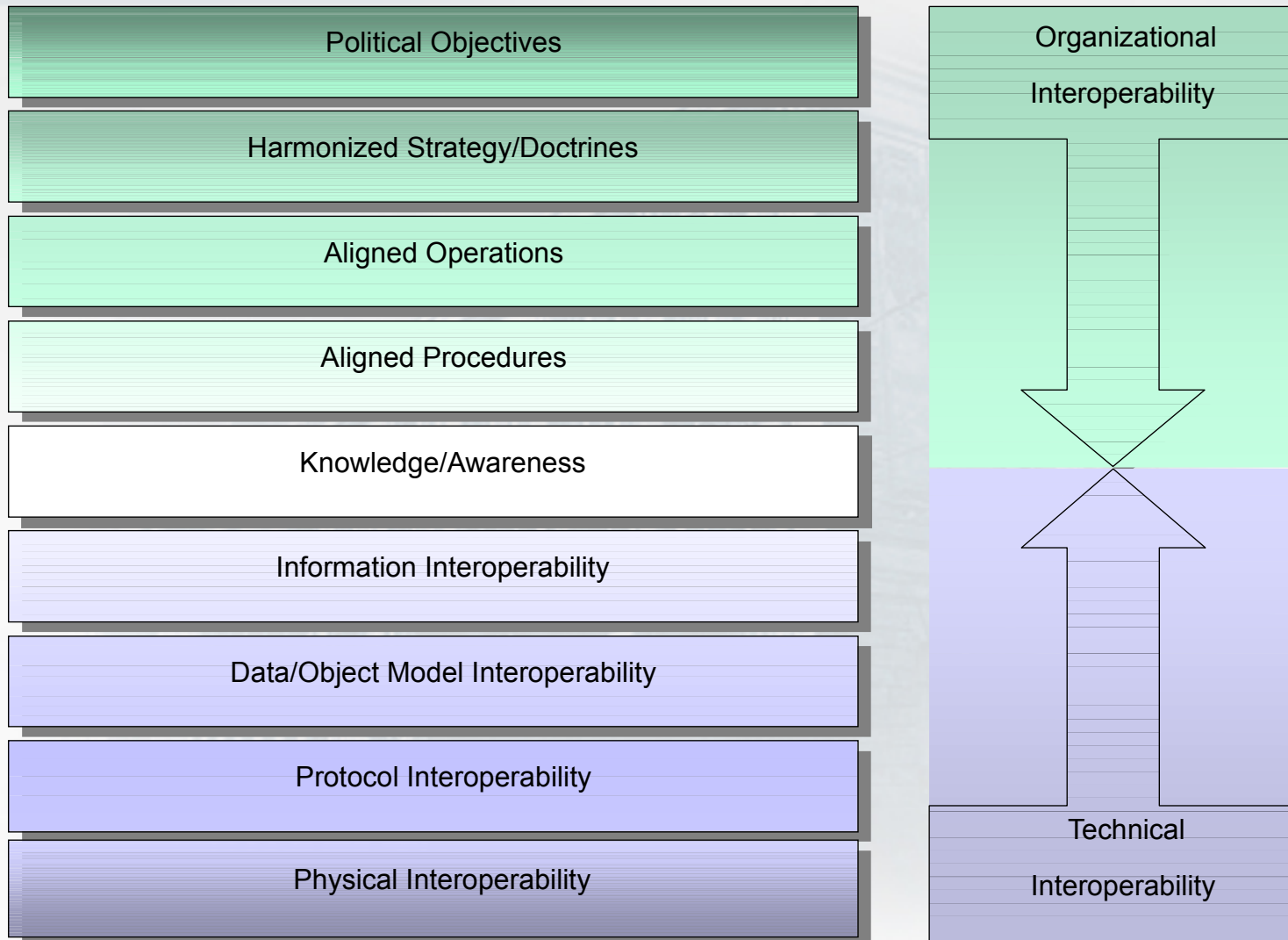


Increasing Capability for Interoperation

Abstraction Layers (Legrand2006)



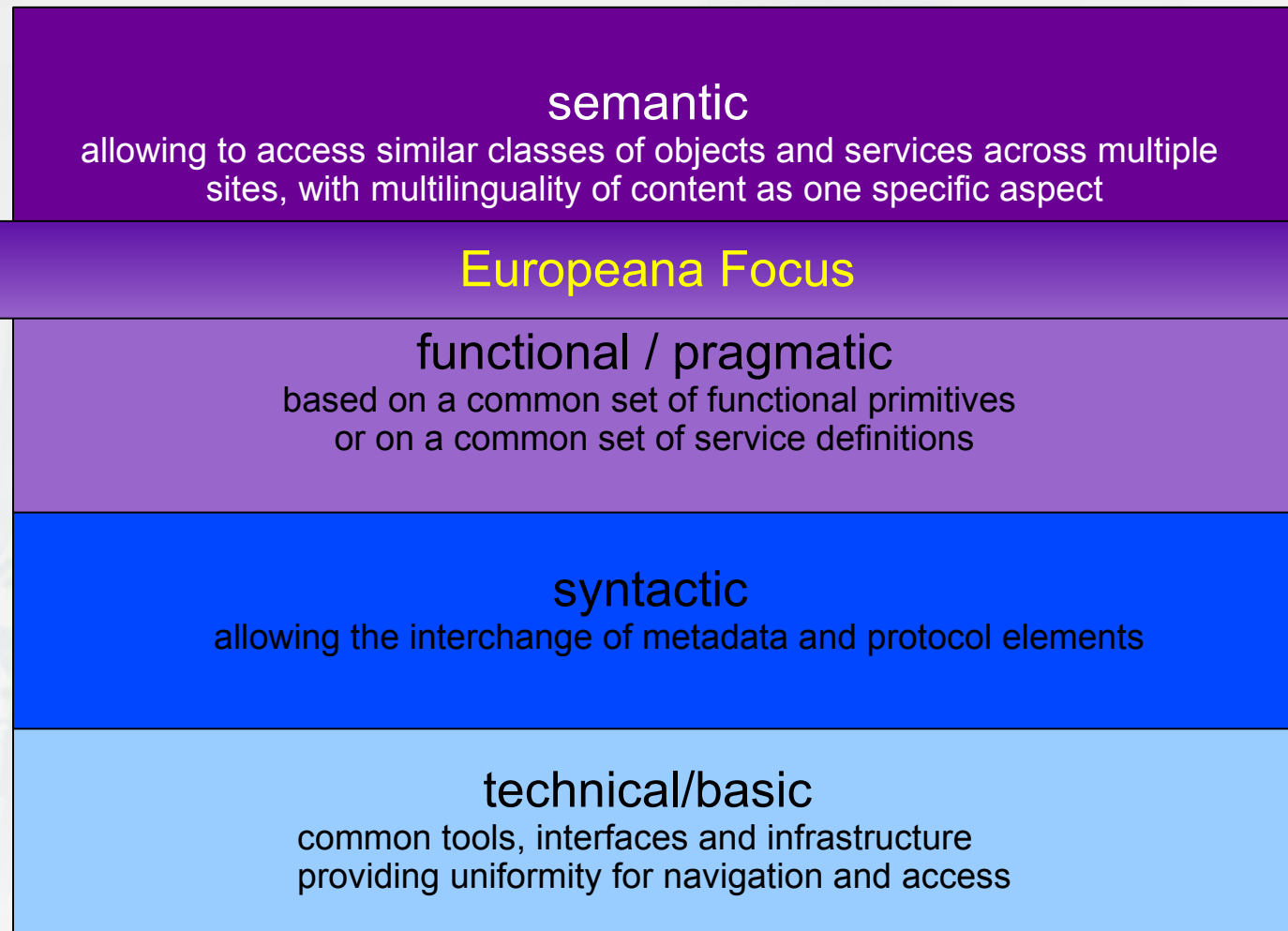
Layers of Interoperability



1

Abstraction Layers

Abstract



Concrete

DL-Interoperability@Firenze 25-09-2008

Entities & Objects



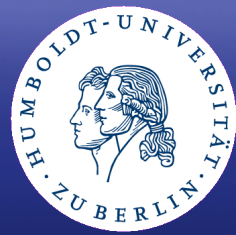
■ Interoperating Entities

- Cultural Heritage Institutions (libraries, museums, archives)
- Digital Libraries,
- Repositories (institutional and other),
- eScience/eLearning platforms or simply
- 'Services'

■ Objects of Inter-Operation

- full content of digital information objects (analogue vs. born digital),
- representations (librarian or other metadata sets),
- surrogates,
- functions,
- services

Functionality and Technology



■ **Functional Perspective of Interoperation**

- Exchange and/or propagation of digital content (OA/Non OA)
- Aggregation of objects into a common content layer (push vs. harvesting / pull)
- interaction with multiple Digital Libraries via unified interfaces
- operations across federated autonomous Digital Libraries (such as searching or meta-analysis for e. g. impact evaluation)
- common service architecture and/or common service definitions or aim at building common portal services.

■ **Interoperability Enabling Technology**

- Z39.50 / SRU+SRW
- harvesting methods based on OAI-PMH
- web service based approaches (SOAP/UDDI)
- Java based API defined in JCR (JSR 170/283)
- Web crawlers & search engines

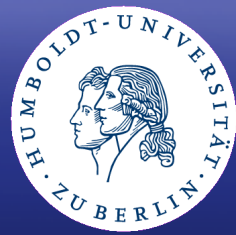
■ Multilingualism

- Multilingual / localised interfaces,
- Multilingual Object Space
 - dynamic query translation,
 - dynamic translation of metadata or
 - dynamic localisation of digital content.

■ Design and Use Perspective

- manager,
- administrator,
- end user as consumer or
- end user as provider of content,
- content aggregator,
- a meta user or a
- policy maker.

The Nasty Bit: Data Quality



- A perfect framework combining
 - solid object modeling
 - well understood functional primitives
 - including authorisation methods
 - as well as using aligned semantic elements
 - and fully multilingual

may still result in a dramatic lack of interoperability:

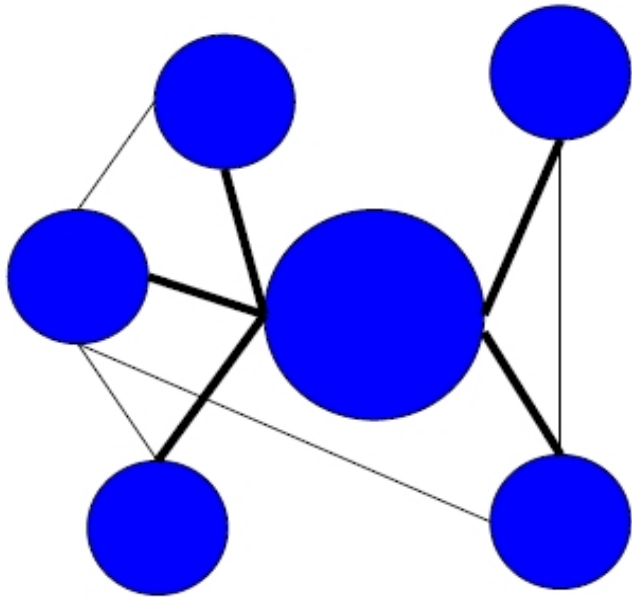
- When operating on 'dirty', heterogeneous data!
- This is a truth both trivial and critical

Interoperability 2.0



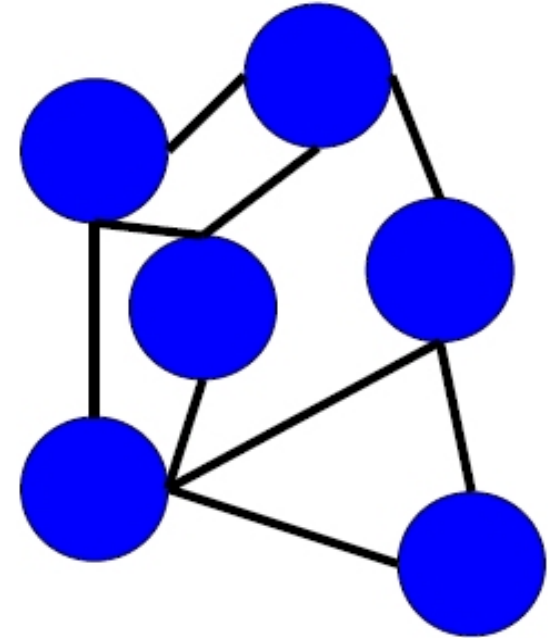
- The 'interoperability' notion has **almost been burnt** by abuse.
- You can even find „**Interoperability 2.0**“ in google (although the disaster is limited to the *IMS Global Learning Consortium Learning Tools Interoperability v2.0 Working Group*)
- Term is used in **surprising contexts**: „Microsoft is committed to solving the real-world interoperability challenges of our customers ...“ => “Document Interoperability Initiative”
<http://www.microsoft.com/interop/principles/default.mspix>
- “While the best example of a communications system based on open standards is the Internet, perhaps the best counter-example lies in the proprietary world of the desktop computing environment, which is dominated by Microsoft’s closed operating system (Windows).” (ECIS, not altogether unbiased!)

Interoperability 2.0



Intraoperability

VS.



Interoperability

- “I think the word “**interoperability**” is being similarly abused. When a single vendor or software provider makes it easier to connect primarily to his or her software, this is more properly called intraoperability.” (Bob Sutor)

Interoperability 1.0



- Re: Motivation (ECIS, slightly modified – DL instead of 'devices'):
 - In today's networked ICT environments, DLs do not function purely on their own, but must interact with other DLs.
 - A DL that cannot interoperate with the other products with which consumers expect it to interoperate is essentially worthless.
 - It is interoperability that drives competition on the merits and innovation.
 - The ability of different DLs to interoperate allows consumers to choose among them.
 - Because consumers can choose among them, DLs must compete with one another, and it is this competition that has driven innovation in the software industry.
- => Interoperability enables **Diversity** and **Competition**.
- Therefore let us keep the concept 'clean' and meaningful.
- This, anyway was the deeper motivation for this seminar.

Further Reading



- (Bekaert 2006) Bekaert, Jeroen and Van de Sompel, Herbert: Augmenting interoperability across scholarly repositories. Report (2006)
<http://msc.mellon.org/Meetings/Interop/FinalReport>
- (Murray 2007) Murray, Robin: Library Systems: Synthesise, Specialise, Mobilise. In: Ariadne 48, July 2006 (<http://www.ariadne.ac.uk/issue48/murray/>)
- (Powell 2005) Powell, Andy: A 'service oriented' view of the JISC Information Environment. University of Bath 2005 (<http://www.ukoln.ac.uk/distributed-systems/jisc-ie/arch/soa/jisc-ie-soa.pdf>)
- (Sompel 2006) Herbert Van de Sompel, Carl Lagoze, Jeroen Bekaert, Xiaoming Liu, Sandy Payette, Simeon Warner. An Interoperable Fabric for Scholarly Value Chains. D-Lib Magazine, 12(10), 2006.
- (Swan 2006) Alma Swan, Chris Awre: Linking UK Repositories: Technical and organisational models to support user-oriented services across institutional and other digital repositories, a JISC Scoping Study Report. 2006.
http://www.jisc.ac.uk/uploaded_documents/Linking_UK_repositories_report.pdf
- (Warneretal 2006) Simeon Warner, Jeroen Bekaert, Carl Lagoze, Xiaoming Liu, Sandy Payette, Herbert Van de Sompel. Pathways: Augmenting interoperability across scholarly repositories. Accepted for IJDL Special Issue on Digital Libraries and eScience. Also available as arXiv:cs/0610031.